



Escuela de Administración

Big Data, un caso vinícola sobre calidad y clima

Trabajo de Grado

Carlos Enrique Rodríguez Coronado

Bogotá

2023



Escuela de Administración

Big Data, un caso vinícola sobre calidad y clima

Trabajo de Grado

Carlos Enrique Rodríguez Coronado

Tutor

Flor Nancy Díaz Piraquive

Título a obtener: Master of Business Administration

Escuela de Administración

19 de septiembre de 2023

Bogotá, Colombia

2023

Contenido

Preliminares	4
Declaración de originalidad y autonomía.....	4
Declaración de exoneración de responsabilidad.....	5
Lista de Figuras	6
Lista de Tablas.....	7
Anexos.....	8
Resumen Ejecutivo.....	9
Palabras clave	10
1. Introducción	12
2. Descripción de la Situación organizacional donde se realizará el proyecto	14
2.1. El Vino	16
2.2. Variables que impactan en la calidad de un vino	16
2.3. La siembra y la agricultura de la vid.....	17
2.4. Cepas	18
2.5. La Industria Vinícola.....	19
2.6. Big Data, ¿qué es y cómo se usa?	22
2.7. ¿Inteligencia Artificial o <i>Big Data</i> ?	23
2.8. Toma De Decisiones Basado En Big Data.....	24
2.9. Impacto del Big Data en la industria.....	25
2.10. Big Data y la industria del vino.....	27
3. Descripción de la situación estudio de caso y/o problemática empresarial	29
4. Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis del estudio de caso y/o solución a la problemática empresarial	32
5. Desarrollo del análisis de datos del estudio del caso	37
6. Plan y recomendaciones de implementación y aplicación	54
7. Hallazgos	56
8. Referencias bibliográficas	58

Preliminares

Declaración de originalidad y autonomía

Declaro bajo gravedad de juramento, que he escrito la presente tesis de maestría por mi propia cuenta, y que, por lo tanto, su contenido es original. Declaro que he indicado clara y precisamente todas las fuentes directas e indirectas de información, y que esta tesis de maestría no ha sido entregada a ninguna otra institución con fines de calificación o publicación.

A handwritten signature in blue ink, consisting of a large, stylized initial 'C' followed by a long horizontal stroke and a smaller 'R' at the end.

Carlos Enrique Rodríguez Coronado
Firmado en Bogotá, D.C. el 19 de septiembre de 2023

Declaración de exoneración de responsabilidad

Declaro que la responsabilidad intelectual del presente trabajo es exclusivamente de su autor. La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él.

A handwritten signature in blue ink, consisting of a large, stylized 'R' followed by a 'C' and a '3', with a long horizontal stroke extending to the right.

Carlos Enrique Rodríguez Coronado
Firmado en Bogotá, D.C. el 19 de septiembre de 2023

Lista de Figuras

Figura 1. OEC: Información sobre comercialización del vino.....	21
Figura 2. Países Exportadores de vino.	21
Figura 3. Estadísticas claves del Big Data.....	27
Figura 4. Variables que afectan la calidad del vino (<i>A.G. Reynolds</i>).....	31
Figura 5. Metodología CRISP-DM	35
Figura 6. Página de inicio de kaggle.com.....	39
Figura 7. Variables Climáticas por Tendencia y Longitud/Latitud de la fuente Global Climate Monitor	40
Figura 8. Resumen de base de datos final con variables principales.....	44
Figura 9. Matriz de Correlación	45
Figura 10. Relación de Calidad de vino (1-100) vs Precios (USD por botella)	46
Figura 11. Relación entre año de producción y calidad del vino	47
Figura 12. Relación entre Calidad del Vino y Temperatura anual promedio de la zona de la cosecha.....	48
Figura 13. Relación entre la Calidad del Vino y Temperatura anual máxima de la zona de la cosecha.....	49
Figura 14. Relación entre la Calidad del Vino y la Temperatura Mínima Anual de la zona de la cosecha.....	50
Figura 15. Relación entre Calidad del Vino y Precipitación Anual de la zona de la cosecha	51
Figura 16. Relación entre Calidad del Vino y Evo Transpiración Anual promedio de la zona de la cosecha.....	52
Figura 17. Relación entre Calidad del Vino y Latitud de la región de la zona de la cosecha	53
Figura 18. Relación entre Calidad del Vino y Longitud de la zona de la cosecha	54

Lista de Tablas

Tabla 1. Tabla de resumen ejecutivo	9
Tabla 2. Tabla comparativa de opciones para abordar la solución	34

Anexos

- Anexo.pdf

Resumen Ejecutivo

Tabla 1. Tabla de resumen ejecutivo

Nombre Del PAE	Big Data: Un caso vinícola sobre calidad y clima
Información Básica Del Proyecto	El presente es un análisis de la industria vinícola, la cual esta sectorizada geográficamente. En este caso, la problemática se centra en poder analizar algunos aspectos generales de la agricultura propia vinícola a través de un análisis de ciencia de datos, para así dar una presentación de posibles orientaciones y disposiciones de como diversificar este tipo de negocio a nivel de sembrado.
Aportes Al Caso De Estudio O La Solución De La Problemática Empresarial	Los potenciales beneficios se dividen en, primero, realizar un análisis combinado de información que permita abstraer información de parámetros sobre la siembra y agricultura vinícola, segundo, mostrar el potencial de la ciencia de datos sobre un análisis de datos para encontrar nuevas opciones a nivel económico, y tercero, realizar una introducción a la importancia de la inteligencia de negocios para la toma de decisiones.

El nombre de este proyecto es *Big Data: Un caso vinícola sobre calidad y clima*, el cual tiene como objetivo realizar un análisis de la industria vinícola, la cual se ha presentado en toda su historia como un proceso y desarrollo sectorizado geográficamente, y donde posiblemente tiene su origen en cuatro zonas del área oeste de Asia (*Where Did Wine Come From?*, s/f). Desde esta referencia del su origen, se encuentra que de por sí ya se encuentra desagregada tanto el cultivo como la producción de vino, llevando esto a nichos en el mercado donde los vinos eran producidos.

De ahí parte este análisis, viendo como los vinos, se concentra en regiones específicas donde pareciera que la calidad depende de estas y también su precio (Corvain, 2022). La pregunta que surge como el planteamiento general de este trabajo es ¿Cómo se

puede determinar cuáles son los mejores rangos de las principales variables climáticas que influyen en la producción de un vino de alta calidad, para así poder tomar una decisión empresarial y de negocio para desplegar un nuevo viñedo? La respuesta no es nada sencilla, ya que por una parte se conocen algunas regiones no convencionales que cultivan uvas para vino (por ejemplo, México, Perú y hasta Villa de Leyva en Colombia) (Webb, 2018), pero no es habitual escuchar que estas regiones se caractericen por un vino de gran calidad. Para este caso, se permite plantear un análisis de algunos aspectos generales de la agricultura propia vinícola a través de la ciencia de datos, para así dar una presentación de posibles orientaciones y disposiciones de como diversificar este tipo de negocio a nivel de sembrado.

Los potenciales resultados esperados son, primero, realizar un análisis combinado de datos recopilados de varios años a nivel mundial que permita abstraer información de parámetros sobre la siembra y agricultura vinícola, segundo, mostrar el potencial de la ciencia de datos para poder inferir y/o concluir información estratégica, relacionada con la toma de decisiones gerenciales y empresariales, y tercero, presentar la importancia de la inteligencia de negocios para la toma de decisiones.

Palabras clave

- *Data Collection and Data Estimation Methodology; Computer Programs*
- *Industry Studies: Primary Products and Construction*
- *Agriculture*
- *Big Data*
- *Inteligencia Artificial*
- *Ciencia de Datos*
- *Vinicultura*

- *Vinos*
- *CRISP-DM*
- *Clima*
- *Vinícola*
- *Calidad*
- *Vid*
- *Cepas*
- *Viñedos*

1. Introducción

Una de las ciencias que más se ha desarrollado y aplicado en los últimos años son la ciencia de dato, la cual “es el estudio de datos con el fin de extraer información significativa para (*¿Qué es la ciencia de datos?*, s/f). Esta está siendo aplicada por múltiples empresas, tal como Amazon, Netflix, IBM, Apple, Coca Cola, entre muchas otras más, las cuales usan la información para conseguir mejores resultados, dependiendo del ámbito. (*10 Most Innovative Data Science Companies in 2023*, s/f)

Por ejemplo, en el caso de Netflix, utiliza la mayor parte de datos para hacer análisis predictivos (recomendaciones, marketing, modelos de precios) (*Predictive Analytics*, s/f), lo que le permite manejar hasta una tasa de 93% retención de clientes, lo cual es sorprendente. Otro ejemplo de uso de datos es la agricultura, donde apoyada por la tecnología del Internet de las cosas (IoT), ha permitido la recolección de datos en tiempo real directamente desde los cultivos mismos, para luego así mejorar las prácticas, control de variables y aumentar la rentabilidad. (*Big Data and Agriculture*, s/f). Es tanta la relevancia que ha cobrado esta materia, que ahora múltiples universidades a nivel mundial ofrecen programas de *Master in Big Data, Analytics, Business Analytics y Data Science*, preparando a los profesionales a un mundo de nuevos tipos de análisis económicos, financieros, directivos y gerenciales. (*43 Best Data Science MBA Programs & Data Analytics MBA*, s/f)

Es por esto por lo que cobra tanta importancia realizar un trabajo de grado basado en la ciencia de datos, en la aplicación de metodologías como CRISP-DM para el análisis de datos (que por cierto es bastante similar al aprendizaje basado en retos), la cual consta de seis fases: Entendimiento del negocio, etapa en la cual se determinan los objetivos del negocio; Entendimiento de los datos, donde se explora y verifican los datos a utilizar; Preparación de los datos, donde los datos son revisados y preparados para la siguiente fase; Modelamiento, fase en la cual se investigan, comparan y construyen modelos para obtener conclusiones,

donde se analiza el mejor modelo según el objetivo; Evaluación, donde se evalúan los resultados de los modelos, se revisan posibles falencias de las anteriores fases; y por último, Despliegue, en el cual se toma los resultados de la evaluación de datos, se plantea un plan de despliegue y monitoreo del mismo. (“Crisp DM Methodology”, s/f)

Es así, que el objetivo de este trabajo es de poder mostrar con un ejemplo práctico y real como con la información que se encuentra disponible, se pueden obtener información y conclusiones que permitan tomar decisiones empresariales, como lo puede ser seleccionar un tipo de negocio, los precios y hasta las zonas donde desplegar dicho negocio, además, de mostrar que la aplicación de la ciencia de datos es primordial para cualquier programa de MBA. Por ejemplo, para la industria vinícola (como se puede encontrar en algunos capítulos de este documento), se proyecta que para 2030 la industria del vino llegue a \$528.3 billones de dólares a un CAGR del 4.2% (anual) entre 2022 y 2030 (Globe Newswire, 2020). Esto permite entender la capacidad de crecimiento de este negocio, además de las oportunidades comerciales, de nuevos emprendimientos y optimización de procesos.

Finalmente, este documento se organiza en capítulos, donde se explica paso a paso la metodología y el análisis de la información, su procesamiento, modelado y conclusiones respectivas para así realizar una implementación a futuro de estos resultados.

2. Descripción de la Situación organizacional donde se realizará el proyecto

“The Big Short” (La gran apuesta) es una película del año 2015, que dramatiza la crisis financiera vivida en Estados Unidos entre el 2007 y 2010. Explica como un grupo de inversores apuesta contra el mercado de bienes raíces, y como a través de sus análisis e investigaciones, encuentran la corrupción de las grandes corporaciones detrás de este mercado. (McKay, 2015)

Resulta curioso como esta historia muestra lo que este grupo de inversores, cada uno por su lado, analizan los procesos, las tendencias y la información, para así apostar contra uno de los más grandes mercados a nivel global y local de Estados Unidos, la finca raíz, las hipotecas y los préstamos bancarios. Muchos de los que han visto la película han pensado, que, de haber tenido acceso a esa información, también hubieran podido apostar contra el sistema y volverse millonarios de un día a otro.

Y es acá donde surge la pregunta: ¿Es posible determinar que puede suceder en el futuro si se encuentra con una basta cantidad de información sobre problemas empresariales, directivos o estratégicos?

La mayoría de las empresas contratan directores con amplia experiencia para que puedan dirigir y desarrollar estrategias, con el fin de tomar las mejores decisiones para el bien y el futuro de esta. Muchos de ellos utilizan lo que se conoce como “*Gut Feeling*” o presentimiento. Es por ejemplo el caso de Bob Lutz, presidente de *Chrysler*, quien en 1988 tomó una decisión que cambió el rumbo de la compañía y la condujo nuevamente a la cima. (Hayashi, 2001)

Existen estudios científicos que indican que este tipo de presentimiento para la toma de decisiones proviene de una habilidad especial de poder detectar patrones sobre situaciones vividas y experiencias pasadas, posiblemente de forma subconsciente, lo cual permite tomar mejores decisiones, lo cual no pueden ser las correctas en todos los casos. (Hayashi, 2001)

Así mismo, existen múltiples artículos que exponen, que, para mejorar la toma de decisiones, se debe mejorar la habilidad de realizar predicciones, como lo es tener mayor certeza del entorno, analizar lo que sucede habitualmente en este tipo de casos, y de tener un pensamiento más “probabilístico”. (Frick, 2018)

En los últimos años se ha desarrollado de manera exponencial el uso de datos, Internet y abrió las puertas para que los individuos puedan obtener información de manera rápida y confiable (aunque también permite la distribución de información y noticias no verídicas), que permiten realizar tareas de forma más rápida y precisa, y a medida que se obtiene más información, se hace crucial realizar un análisis de esta información para poder tomar mejores decisiones.

El caso de estudio de ese documento se centra en lo descrito anteriormente, un ejemplo de cómo con el correcto análisis de la información disponible, se puede tomar mejores decisiones, lo que nos permita ser mejores administradores, gerentes y directivos en nuestras empresas y emprendimientos. De hecho, uno de los puntos expuestos por Marco Pastrana en su clase de Administración de Empresas (MBA, Rosario GSB, Universidad del Rosario) planteó que vislumbra en el futuro de las compañías, un método donde las gerencias empresariales estén soportadas por procesos de *Big Data*, el cual permitan que la empresa tome decisiones más acertadas (Pastrana De La Cruz, 2019)

Sin embargo, ¿Cómo se puede tomar mejores decisiones a través de gran cantidad de información?

Es así como se plantean los objetivos de estudio y análisis de este documento, donde se describe la metodología CRISP-DM, con la cual se utilizan tecnologías como *Big Data Analytics & Science* para realizar análisis y conclusiones de grandes cantidades de datos, y así, tomar mejores decisiones de inversión, económicas, estratégicas y administrativas.

Este documento no tiene como objetivo describir cómo se realiza el análisis de *Big Data*, o entrar en detalle de cuestiones técnicas específicas de cómo desarrollar programas de análisis de datos, proyecciones estadísticas ni cálculos teóricos, sino el de mostrar un desarrollo de análisis sobre un ejemplo práctico, el cual permite deducir tendencias y tomar decisiones de inversión, siendo el caso en específico, el tema vinícola.

2.1.El Vino

Tal como lo menciona el libro “El mundo del vino”, de la editorial Larousse, la mayoría de los historiadores concuerdan que el origen del vino no es el Mediterráneo (aunque sea el lugar donde más se produce vino actualmente), sino del área Caucásica, aunque también hay indicios de primeros cultivos en otras tierras asiáticas o en el norte de África. (Yanet Acosta et al, 2013)

Este libro, que recopila información sumamente interesante con respecto al vino, la cultura, las cepas, regiones, viticultura, entre otros, explica que “la vid es un cultivo adaptado a los climas templados y que, en general, encuentra problemas en los lugares gélidos y en aquellos otros que desprenden excesivo calor y humedad, ya que requiere de veranos cálidos e inviernos fríos y su ciclo es aún mejor cuando existe diferencia térmica importante entre las temperaturas diurnas y las nocturnas”. Esto permite inferir que hay regiones donde la vid no pueda ser cosechada para que su proceso entregue un vino de calidad aceptable.

2.2.Variables que impactan en la calidad de un vino

En el capítulo “Del subsuelo a la mesa” del libro “El mundo del vino” (Yanet Acosta et al, 2013), se comenta sobre múltiples factores que pueden afectar o colaborar en la producción de un vino de buena calidad. Más allá de describir las diferentes variables, se hace mención de que esta es una “Ciencia al aire libre”, ya que contiene múltiples focos de análisis, como lo son:

Geología: Un estudio del subsuelo permite conocer la profundidad cultivable de la vid, la capacidad de retención de agua y la composición mineral del suelo.

Edafología y Ecología: Permite encontrar particularidades sobre la superficie del terreno, los nutrientes orgánicos y organismos del ambiente.

Meteorología: Se obtiene la información climática del área, con datos como la precipitación promedio anual, exposición solar, temperaturas máximas, mínimas y promedio, vientos, riesgos climáticos y detalles que permitan pronosticar el clima para la cosecha. Es en este punto que se centra este trabajo, analizando información de miles de calificaciones de vinos, asociándolos a sus zonas de cosecha y variables climáticas dependiendo del año.

Biología, Fisiología Vegetal y Ampelografía: Estas áreas ayudan a determinar cuáles son las cepas que podrían ser más acordes para el clima y tipo de suelo de la zona y el terreno.

Estas áreas permiten preparar y entender al viticultor las principales variables que pueden ayudar a su cosecha, o en el peor de los casos, a afectar toda una cosecha de vid.

2.3. La siembra y la agricultura de la vid

El proceso manual para cosechar la vid, en la mayoría de los casos, consta de los siguientes pasos, según el Blog de Jean Leon sobre el ciclo de la vid. (Leon, 2017)

Plantación y Poda: Consta del proceso de plantar la vid en la cosecha y de permitir su crecimiento. Así mismo, en el ciclo completo de la cosecha en esta época de invierno, la viña se poda para dejarla solo como un tronco sin floración. Esto permite que soporte las bajas temperaturas y condiciones atmosféricas fuertes de la temporada fría.

Inflorescencia: En esta fase comienza la parte más intensa de la cosecha, donde luego de un estado de reposo, comienza el movimiento de savia en las primeras brotaciones, para luego limpiar las hierbas, atar los sarmientos y escoger cuales serán las plantas que se dejarán crecer para obtener nuevos frutos.

Floración: Con la estación de la primavera, aparecen los primeros brotes de la vid y donde germinan las primeras viñas con algunas hojas en la planta. En este periodo, el

viticultor termina de podar la viña y de realizar procesos de protección de plagas y enfermedades.

Cuajado: A medida que se acerca el verano, se abre la flor para ser fecundada. En esta etapa de dos semanas es sumamente importante que el tiempo sea calmado (ya que una fuerte tormenta puede llegar a afectar y dañar la cosecha), por lo que el control de la cosecha y el permanente proceso de control de plagas y enfermedades es de suma importancia.

Despunte y Envero: En la estación de verano es cuando el fruto madura, por lo cual se debe tener cuidado que el fruto no sea maltratado, y que se permita que madure y supervisando un buen crecimiento. Se conoce como “envero” el momento en que la uva se torna de un color rojizo para los tintos, o amarillento para los blancos. También en algunos casos se requiere un proceso de despunte de los frutos para controlar los frutos que puedan presentar deficiencias en su maduración.

Vendimia: Es la etapa donde se prueba la uva para evaluar sus niveles de acidez y azúcar, esto permite entender el mejor momento para poder recolectar las uvas (las cuales se recomiendan coleccionar de forma manual para no afectar el fruto ni la vid). A continuación, son transportadas hasta las tolvas (recipientes para la recepción de las uvas), para luego continuar con el proceso de despallado, prensado, fermentación, envejecimiento y tratamiento antes de pasar al proceso de embotellado.

Caída de la hoja: En esta fase, con la llegada de la estación de otoño, se comienza el proceso de podar nuevamente la vid, para así ir preparándola para el otoño. También, como se acerca la etapa más fuerte la para la vid, se recomienda nutrir la tierra con abono para así poder asegurar la calidad de los frutos futuros.

2.4. Cepas

A este momento, existen más de 3'000 variedades de cepas de uva para hacer vino (*Vinos y Sabores*, s/f), y alrededor de 6'000 variedades de uvas registradas (Yanet Acosta et al, 2013), esto es lo que permite que existan múltiples variedades de sabores de vinos, y esto,

combinándolo con el tipo de suelo, terreno, clima, procesos de cosecha y recolección, fermentación y bodegaje, hacen que el mundo del vino tenga una diversidad “infinita” de sabores y calidades.

Entre las principales cepas, se pueden encontrar (Yanet Acosta et al, 2013):

- *Albariño*: Uva blanca con sabores frutales y florales
- *Cabernet Sauvignon*: Uva tinta de sabor tónico
- *Chardonnay*: Uva blanca de sabor suave
- *Chenin Blanc*: Uva blanca de buena calidad
- *Garnacha*: Uva tinta de gusto cálido
- *Gewürztraminer*: Uva tinta con sabor a fruta pronunciado
- *Macabeo*: Uva blanca de aromas variados
- *Malbec*: Uva tinta de suaves taninos
- *Merlot*: Uva tinta de aromas elegantes
- *Moscatel*: Uva blanca o rojiza que produce vinos variados
- *Nebbiolo*: Uva tinta con buen equilibrio de acidez y aromas
- *Palomino*: Uva blanca con sabor amargo
- *Pedro Ximénez*: Uva blanca de gusto a frutas
- *Pinot Noir*: Uva tinta de sabor afrutado
- *Riesling*: Uva blanca de sabor aromático
- *Syrah*: Uva tinta suave y aromático
- *Tannat*: Uva tinta con gran presencia de taninos
- *Tempranillo*: Uva tinta que permite su mezcla con otras variedades
- *Verdeo*: Uva blanca con aromas y sabores complejos
- *Zinfandel*: Uva tinta de buen balance de textura, cuerpo y acidez

Cada cepa tiene un proceso diferente, por lo cual es importante tener en cuenta el tiempo de cosecha y recolección.

2.5. La Industria Vinícola

La industria vinícola se ha mantenido sin muchos cambios a lo largo de los últimos años, sin embargo, solamente en Estados Unidos, este mercado ha pasado de \$30'000 millones de dólares en 2002 a más de \$60'000 millones de dólares al 2019, convirtiéndolo en la más grande a nivel mundial. Las casas vinícolas han crecido en un 50% en la última década, y donde se ha redefinido que es un gran vino, como se maneja la lealtad de sus consumidores apasionados y mantener precios extraordinarios. (“What the Wine Industry Understands About Connecting with Consumers”, 2019)

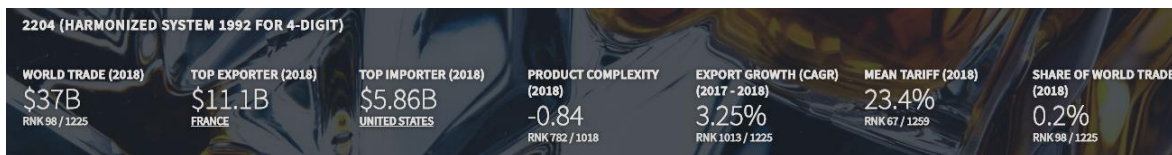
Aún con el impacto del coronavirus y las cuarentenas a nivel mundial, se estima que la industria vinícola seguirá creciendo. En el reporte “Global Wine Industry” (reportlinker.com), se prevé un crecimiento de \$60'000 millones de dólares para 2027, con una Tasa de Crecimiento Anual Compuesto, TCAC (CAGR, Compound Annual Growth Rate) del 4,2% entre el periodo 2020 a 2027. Así mismo, economías como China tienen una proyección de TCAC del 7,4%. (Globe Newswire, 2020)

Según la OEC (The Observatory of Economic Complexity, oec.world), en 2018, el vino fue el 98° producto más comercializado a nivel mundial, y donde las acciones representan un 0,2% del total del mundo. (*Wine HS*, 2018)

En cifras generales, el mercado del vino en 2018 obtuvo:

- Comercialización por más de \$37 miles de millones de dólares (producto 98° de 1225)
- El mayor exportador es Francia con \$11,1 miles de millones de dólares
- El mayor importador es Estados Unidos con \$5,86 miles de millones de dólares
- El crecimiento de exportación – TCAC es del 3,25% (2017-2018)

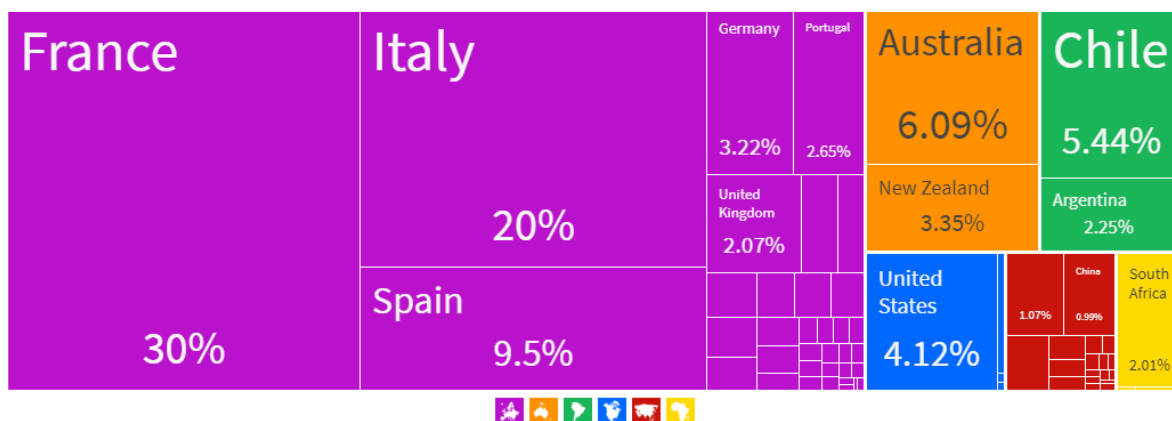
Figura 1. OEC: Información sobre comercialización del vino.



Nota. Adaptado de OEC: *Wine of fresh grapes*, por OEC World, 2017, oec.world (<https://oec.world/en/profile/hs/wine>)

Entre los países que más exportan vino, se encuentran Francia, Italia, España, Australia, entre otros, donde los referentes del continente se encuentran Estados Unidos (con sus cultivos conocidos de California), Chile (con cultivos alrededor de la Ruta del Vino, en Santa Cruz) y Argentina (regiones de Mendoza y San Juan).

Figura 2. Países Exportadores de vino.



Nota. Adaptado de OEC: *Which countries export Wine?*, por OEC World, 2018, oec.world (https://oec.world/en/visualize/tree_map/hs92/export/show/all/42204/2018/).

En el ejemplo de Francia, el vino contempla el 1,97% del total de exportaciones del país (2018).

Más allá de la comercialización del vino, existen muchos negocios y carreras alrededor de este mercado. Entre ellas se referencian algunas como:

- Enólogos
- Cultivadores de vino

- Logística y Operaciones
- Ingenieros Ambientales
- *Sommeliers*
- Distribuidores
- Ventas y Administración
- Diseñadores gráficos
- Marketing
- Educación

Debido a su continuo crecimiento e importancia a nivel regional y mundial que se hace importante el análisis de este tipo de mercado, y poder entender cuáles son las variables que impactan la calidad y comercialización de esta industria.

2.6. Big Data, ¿qué es y cómo se usa?

¿Qué es *Big Data*? Según Google, lo definen como una gran cantidad de datos disponible para las organizaciones, el cual, debido a su complejidad y volumen, no es fácilmente manejado o analizado por algunas herramientas de inteligencia de negocios (*What Is Big Data?*, s/f). Sus características principales son:

Volumen: Es la característica principal, debido a la cantidad de datos recolectados de múltiples fuentes y dispositivos.

Variiedad: Se refiere a los formatos en que viene la información, para subsecuentemente ser clasificada

Velocidad: Hace referencia a la velocidad que un *dataset* puede ser obtenido, accedido y procesado.

Variabilidad: Corresponde a que la información cambia constantemente, por lo que antes de que la información sea procesada, debe ser debidamente comprendida.

A medida que el *Big Data* ha emergido, así mismo han evolucionado los modelos computacionales y de procesamiento para hacer frente a este tipo de información, permitiendo que las compañías puedan obtener perspectivas y hallazgos, que permitan entender las áreas que afectan la empresa, desde condiciones del mercado y comportamientos de compra de los clientes hasta procesos del negocio.

En el mercado se encuentran múltiples herramientas para trabajar con *Big Data*, muchas de ellas dependen de las 4 características mencionadas anteriormente. Entre las herramientas más conocidas están (Vaughan, 2023):

BigQuery (Google): Plataforma en la nube de alta escalabilidad diseñada para la agilidad empresarial.

Hadoop (Apache): Permite el procesamiento distribuido de *datasets* a través de múltiples clústers de computadores usando modelos de programación simple.

Snowflake: Plataforma en la nube con procesamiento SQL simplificado

Tableau: Herramienta que permite con un enfoque en la visualización de datos

PowerBI: Contiene una comunidad muy grande de usuarios, que permite una integración con el ecosistema de Microsoft y posee excelentes características de seguridad

Para análisis de datos se pueden utilizar herramientas como **R**, el cual es un lenguaje estadístico usado para el análisis y representación física de los datos, o como **Python**, el cual está orientado a la aplicación de modelos de *machine learning* y *deep learning*. (*Python or R for Data Analysis*, 2023)

2.7. ¿Inteligencia Artificial o *Big Data*?

Aunque ambos términos tienen cierta relación, ya que la Inteligencia Artificial hace uso de *Big Data*, no son lo mismo.

En el caso de *Big Data*, las empresas usan la gran cantidad de información que tienen para así analizarlas y a partir de ellas tomar decisiones estratégicas. Es acá donde los hallazgos de la información procesada resultan útiles para tomar dichas decisiones.

En el caso de Inteligencia Artificial (IA), esta da sentido a los datos, donde a través de modelos de aprendizaje va aprendiendo ciertos resultados, aunque también permite seguir aprendiendo de nueva información para así ir mejorando los resultados o predicciones.

Para el objetivo del proyecto se enfoca en *Big Data*, sin embargo, se podría ampliar en el futuro sobre la IA, para poder hacer predicciones de precios basándose en la información de los *datasets* originales. También podrían hacerse modelos predictivos de calidad basándose en la ubicación.

2.8.Toma De Decisiones Basado En Big Data

Uno de los puntos presentados en la clase de Administración de Empresas (MBA, Rosario GSB, Universidad del Rosario) por el profesor Marco Pastrana, expone que la innovación es el *asset* más importante para las empresas en la actualidad, las empresas que no estén dispuestas a invertir en innovación no podrán liderar el mercado y se arriesgan a que estas desaparezcan del mercado. (Pastrana De La Cruz, 2019)

Así mismo, otro de los puntos presentados en la clase Dirección de Marketing por la profesora Lina María Echeverri (MBA, Rosario GSB, Universidad del Rosario), el ciclo de vida de una empresa es cada vez más corto, y donde se hace necesario que las empresas entiendan y apliquen estrategias de marketing, conociendo cada vez más a su cliente, entendiendo a profundidad sus hábitos y así poder ganar su lealtad. (Echeverri, 2019)

Estos análisis académicos concuerdan con el estudio “*The Deciding Factor: Big Data & Decision Making*”, realizado por la consultora *Capgemini* para *The Economist*, el *Big Data* representa un cambio fundamental en la toma de decisiones a nivel empresarial, donde comúnmente las compañías estaban acostumbradas a analizar ciertos datos internos,

pero en la actualidad están haciendo uso de fuentes externas y datos en tiempo real para generar nuevas oportunidades y diferenciarse en el mercado.

Uno de los muchos beneficios del *Big Data* como herramienta innovadora es el de ganar nuevas percepciones sobre los clientes, mercados, cadenas de operación y logística; es decir, tener una nueva perspectiva del negocio que antes no se podía tener para poder marcar una estrategia en el mercado. (“The Deciding Factor”, 2012)

2.9. Impacto del Big Data en la industria

Según un artículo “7 real-world examples of how brands are using big data analytics” (Tudor, 2021), en los últimos 2 años las empresas a nivel mundial gastan más de 180’000 millones de dólares por año en análisis de *Big Data*.

Aunque estas cifras son altas, aún existen muchas compañías que aun no usan esto como una oportunidad para comenzar a invertir en *Big Data*. A continuación, se presentan algunos casos reales con información de como el Big Data ha ayudado a empresas a reducir sus costos, mejorar sus ganancias u optimizar sus procesos.

Universidad Purdue (Indiana, USA): En 2007, esta universidad lanzó un sistema llamado *Signals*, el cual fue diseñado para ayudar a predecir problemas de comportamiento y aprendizaje por materia. Como resultado, los estudiantes que tomaban dos o más asignaturas con *Signals*, tuvieron un 21% menos de probabilidades de desertar. (Hillier, 2021)

UPS: A través del seguimiento del clima y sensores en los camiones, UPS aprendió las rutas más rápidas para sus conductores. Más allá que este es un hallazgo interesante, encontraron que girar a la izquierda en el tráfico hacía que gastaran altos niveles de combustible. Como resultado, implementaron la política “no girar a la izquierda”. Con esto, la compañía asegura que ahora usan menos de 10 millones de galones de combustible al año, y emiten 20’000 toneladas menos de dióxido de carbono (Hillier, 2021). Si se toma un precio referencial de \$4’00 dólares por galón, se pueden estimar ahorros de hasta 40 millones de

dólares al año, sin contar con menores gastos operativos como lo pueden tiempos de reaprovisionamiento de combustible.

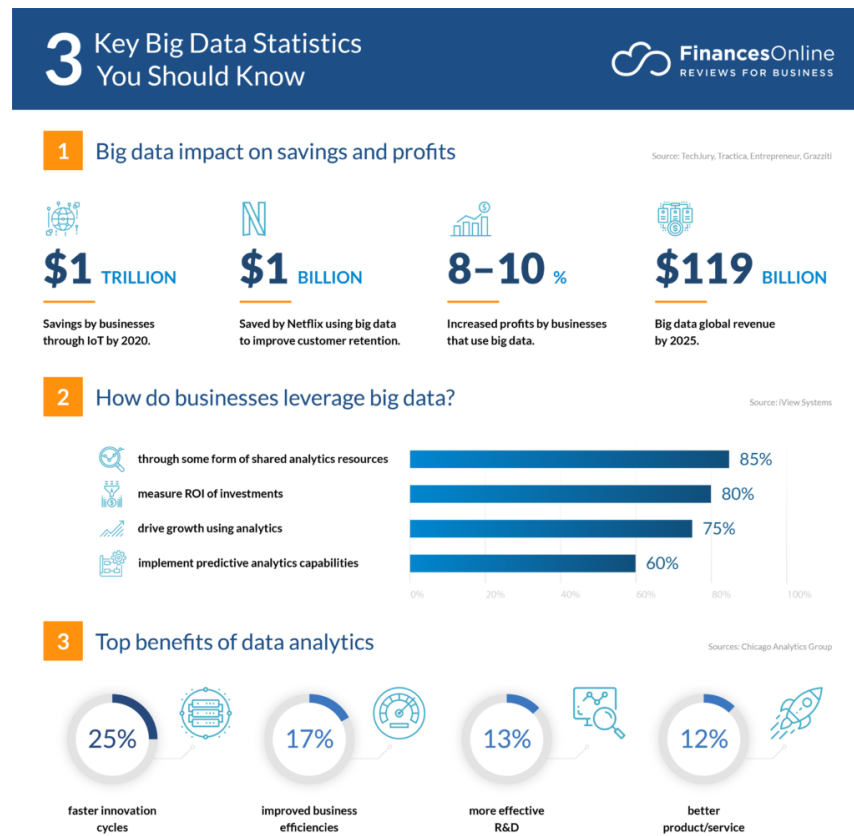
Marriott: Desde el 2014, en sus hoteles *Starwood*, han utilizado el precio dinámico, basándose en una variedad de factores tales como situaciones de la economía local y global, clima, disponibilidad y comportamiento de las reservaciones, cancelaciones, entre muchos otros. Con esto, han podido mejorar en un 5% sus ganancias por cuarto en solo 12 meses. También han ido más allá donde siguen a famosos músicos que tocan en el Madison Square Garden, para así ajustar sus costos en los hoteles próximos a la arena. En el análisis financiero esto equivale a un costo por cuarto por año de \$13'500 dólares, \$1'125 dólares por mes y a \$37 dólares por día. (IntelligentHQ, 2020)

Walmart: Lanzó su programa “*Walmart Lab*” para empoderar su *Big Data* e impulsar su crecimiento. A partir de esto creó un programa de fidelización y retención, llamado *Walmart+*, donde usa modelos de análisis predictivos para entender a su clientela y mejorar sus programas para diferentes segmentos. Este programa cobra \$98 dólares anuales por cliente y ofrece múltiples beneficios. Como resultado, la compañía ha triplicado sus ventas en línea de \$21'900 millones de dólares en 2019 a \$64'620 millones de dólares en 2021. (Pandey, 2022)

Es así como con solo algunos ejemplos se encuentra los grandes beneficios que se tienen al aplicar *Big Data* en las empresas. Basándose en el reporte de *financesonline.com* (Jay, 2019), se obtiene información de algunos ejemplos de los beneficios de *Big Data*, como lo son:

- Netflix ahorra \$1'000 millones de dólares por año en retención de clientes
- 8% a 10% de incremento de ganancias en los negocios que usan Big Data
- 25% de mejoras de tiempos en ciclos de innovación
- 17% de optimización en productividad
- 13% de efectividad en Investigación y Desarrollo

Figura 3. Estadísticas claves del Big Data



Nota. Adaptado de *3 Key Big Data Statistics You Should Know*, por FinancesOnline, 2019, FinancesOnline.com (<https://financesonline.com/big-data-statistics/>)

2.10. Big Data y la industria del vino

En la industria del vino y la cosecha de la vid también se encuentran ejemplos de cómo el Big Data ha ayudado a cambiar las prácticas que se han usado por cientos de años con relación al vino, por ejemplo:

Mejorando la irrigación e implementando gestiones sostenibles del agua: El vino es una fruta la cual su calidad depende en gran parte en el nivel de humedad, así mismo, existen regiones que son muy buenas para cosechar la vid, pero que pueden ser propensas a sequías o picos de calor altos. Es por esto por lo que granjeros y cosechadores están

apoyándose en la analítica de datos para contrarrestar algunos de estos problemas. Es importante también resaltar que esto está impulsado por el Internet de las Cosas (IOT), ya que permite también poder tomar información en tiempo real para así tomar decisiones que garanticen una buena cosecha. El viñedo *Palmaz* en el valle de Napa es uno de los que ha invertido por más de diez años en sistemas de datos e inteligencia artificial para así medir, procesar y visualizar millones de datos para así producir un producto de gran calidad. (Dutta, 2004)

Según el artículo de Forbes “*How Big Data Can Predict The Wine Of The Century*” escrito por James Marland, (Marland, 2014); ¿Puede el Big Data ayudar a encontrar una buena botella de vino?, pues este planteamiento es algo más común de lo que se cree. De acuerdo con el libro “*Super Crunchers: How Anything Can Be Predicted*” escrito por Ian Ayres, la respuesta es que sí. Aunque sea costumbre que para saber la calidad de un vino se contrate a un catador que pueda probarlo y dar sus comentarios sobre la complejidad de sabores, esto ha cambiado en los últimos años. En este caso, Orleu Ashenfelter, un economista planteó una fórmula para determinar la calidad de un vino, usando la cantidad de lluvia en invierno, más la temperatura promedio menos la lluvia en el periodo de recolección, incluyendo ciertos factores en la ecuación. Como era de esperarse, los expertos lo criticaron, sin embargo, Orleu hizo bastante dinero con su fórmula.

Esto, junto con otros casos reales sobre Big Data e Inteligencia Artificial, presenta nuevas opciones para abordar problemas, y analizar procesos y situaciones que anteriormente solo era posible con la intuición o con expertos en la materia.

3. Descripción de la situación estudio de caso y/o problemática empresarial

Si se investiga en Internet sobre *Big Data*, *datasets* y *wine quality*, se encuentran múltiples estudios y artículos sobre calidad del vino. Estos se centran en análisis fisicoquímicos y químicos, relacionando la calidad con algunos compuestos, por ejemplo: acidez, azúcar residual, sulfatos, densidad, pH, y alcohol, entre otros. (Nguyen, 2020)

Este tipo de estudios permite crear un modelo predictivo para la industria, basado en la producción de un vino y aplicando un análisis para poder estimar con cual calidad puede ser una cosecha. Esto puede estar limitado a ciertas variables químicas o, como el ejemplo referenciado, a un *dataset* restringido, donde los datos de vinos de alta calidad con respecto a las variables fisicoquímicas y químicas es restringido.

Tal como lo menciona el artículo “Big Data Analytics Changing the Centruies-Old Wine Industry” (*Big Data Analytics Changing the Centuries-Old Wine Industry*, s/f), los viñedos están aplicando técnicas avanzadas como analítica de datos para poder superar algunos desafíos durante sus procesos, desde el uso de estas herramientas en la parte de marketing y ventas, hasta el uso del internet de las cosas para monitorear y supervisar en tiempo real las cosechas.

Se encuentran estudios como el “*Historical long-term cultivar×climate suitability data to inform viticultural adaptation to climate change*” presentado en la revista *Nature* en 2022, el cual justamente hace un análisis a través del tiempo sobre los cultivos vinícolas, su relación con el clima y sobre la adaptación a los cambios climáticos. En el se expresa la compleja relación que existe entre el cultivo y los índices bioclimáticos, los cuales son importantes para la uva, para así poder evaluar y mitigar los efectos de los cambios climáticos en la calidad de la uva. Este análisis lo llevaron a cabo en la región de Napa Valley, analizando los cambios climáticos de los últimos años con respecto a un historial de más de 100 años, basándose en variables climatológica como temperatura máxima y mínima del aire, temperatura promedio, acidez absoluta del suelo y pH. (Bai et al., 2022). Lo más interesante de esto es la importancia que está tomando en los años recientes el uso y análisis de la ciencia

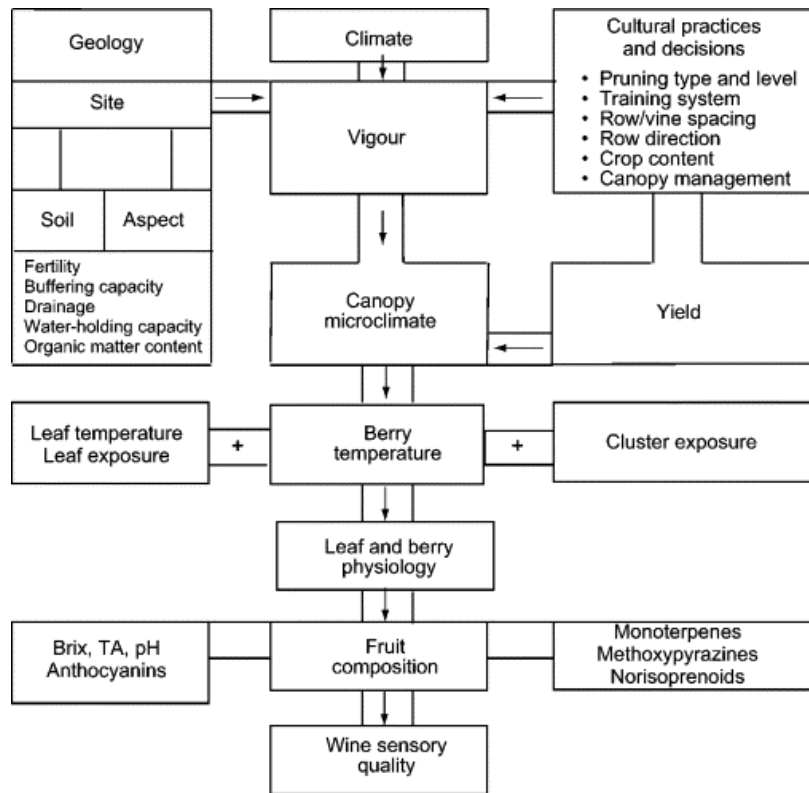
de datos para poder aplicar en casos de uso actuales y de impactos ambientales a nivel global; de hecho, hay artículos de años anteriores (Jones et al., 2005), en el cual realizan un análisis comparativo de la calidad del vino con respecto a los cambios climáticos, donde en una de sus conclusiones es que el cambio climático puede llegar a impactar en pasar valores límites que afecten la calidad del vino. En este estudio es interesante analizar las regiones centroides donde se produce el vino, las cuales se ven reflejadas en el estudio de este trabajo. Otro trabajo similar es el de “*Global warming and wine quality: are we close to the tipping point?*”, realizado en 2021 (Gambetta & Kurtural, 2021), donde hacen análisis específicos de regiones y cepas donde concluyen que las métricas indican que se puede estar llegando a unos valores climáticos que puedan afectar la calidad del vino.

También se encuentran otro tipo de investigaciones como el de *HAL Open Science* del año 2020, donde se analizan temas como la radiación ultravioleta y cambios de las fechas de cosechas del vino para adaptarse al cambio climático de los últimos años, y donde concluyen que los cosechadores deben implementar prácticas y estrategias para poder controlar dichos cambios. (Cornelis van Leeuwen, Philippe Darriet, 2020)

También se han encontrado artículos que hacen mención en la importancia del suelo, de la humedad y de las temperaturas, con análisis principalmente relacionados a una apreciación subjetiva o a una caracterización generalizada por las áreas de cosecha de los vinos. (Jackson, 2017)

Para el objetivo de este estudio, se necesitan entender algunas de las variables que pueden influir en la calidad de un vino. Para esto, se hace referencia a algunos artículos y libros, Entre estos se destaca el libro *Managing Wine Quality: Viticulture and Wine Quality*, 2010, de A.G. Reynolds, donde se destacan las múltiples interacciones de los componentes, el clima, la temperatura, el sol y otros sobre la calidad del vino, tal como se puede apreciar en la siguiente figura.

Figura 4. Variables que afectan la calidad del vino (A.G. Reynolds)



Nota. Adaptado del libro *Managing Wine Quality: Viticulture and Wine Quality*, por Andrew G. Reynolds, 2010, Elsevier Publishers

Entre las principales variables que es del interés sobre el clima, se encuentran:

- Clima
- Microclima del sembrado (humedad, variación de temperaturas)
- Temperatura / Exposición
- Fisiología de las hojas y el fruto
- Componentes y prácticas de sembrado

Es así, que se propone realizar el respectivo análisis de calidad de vinos, obteniendo información por aparte, uno desde el punto de vista de calidad de vino, reseñas de múltiples contribuidores a nivel global, combinándolas con información climática de múltiples puntos geográficos, para así poder obtener a través de una segmentación de datos,

conclusiones sobre las mejores regiones para cosechar vinos, o valores que permitan ser controlados en una cosecha para mejorar las probabilidades de tener un vino de buena calidad.

4. Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis del estudio de caso y/o solución a la problemática empresarial

En la actualidad existen varias opciones para poder abordar la solución al problema planteado. Para esto se describen las siguientes alternativas:

a. Análisis de Mercado a través de viñedos

“Se llama análisis de mercado al proceso de recopilar información necesaria sobre el sector en el que tu empresa participa o al que quiere integrarse. En él se estudian las condiciones actuales del comercio (a nivel macroeconómico y microeconómico) con el fin de tomar decisiones oportunas y rentables para tu negocio” (Santos, 2023). Esta es una de las opciones para abordar este problema. Entre las ventajas está el comprender y entender no solo la recopilación de datos propuestos, sino otras variables y aprendizajes que tienen los viñedos a nivel de agricultura y cosecha de la vid, así como consejos y prácticas particulares. En las desventajas se encuentra que en Colombia hay muy pocos viñedos (algunos en Villa de Leyva) y para poder hacer un estudio de mercado enriquecido e interesante sería necesario viajar a diferentes zonas de Estados Unidos, Chile, Argentina, España, Francia, Italia, Alemania, Australia, entre otros lugares.

b. Conocimiento directo de *sommeliers*, enólogos, catadores de vino

El recabar información directa de expertos también sería una propuesta similar a un análisis de mercado. Esta opción es interesante ya que la ventaja es que los expertos pueden dar su opinión de que variables y rangos puede ser los mejores para un buen vino, sino que también se enriquecería por la información de la calidad de los vinos. Lamentablemente para poder tener un muestreo importante e interesante sería necesario

contactar a muchos de estos expertos. Sería interesante como un paso a futuro hacer consultas a algunos de ellos sobre los resultados del análisis presentado en este trabajo.

c. Información de agricultores y cosechadores

Esta opción es interesante por el hecho de años de experiencia con particularidades climatológicas y experiencia en todo el proceso. De hecho, existen teorías que las fases lunares pueden afectar el desarrollo, crecimiento y calidad de los cultivos. (*La influencia de la luna en la agricultura - Colombia Verde, 2023*). La ventaja es que en Colombia se pueden encontrar con muchos agricultores en pueblos y fincas con este tipo de experiencia, enriqueciendo la información, sin embargo, tendría como desventaja que serían conocedores de la agricultura en Colombia (la cual no cuenta con las estaciones que requiere la vid) y que tengan experiencia sobre cultivos diferentes a la vid, sobre todo para la producción de vino.

d. Análisis con *Big Data*

La opción de utilizar *Big Data* como opción de análisis para este problema es importante, ya que se cuentan con fuentes confiables de bases de datos a nivel de calidad del vino de múltiples catadores, con información enriquecida como precios por botella, vinos de diferentes viñedos, cepas y años; también permite conectar esta información con información climatológica de diferentes estaciones de monitoreo, permitiendo así una gran cantidad de información para poder analizar y obtener hallazgos interesantes. La desventaja es el desarrollo técnico de este análisis que requiere conocimientos sobre la ciencia de datos, programación y conocimientos estadísticos.

e. Análisis con Inteligencia Artificial

Utilizar Inteligencia Artificial estaría enfocado a crear un algoritmo de análisis que pueda permitir el análisis de variables para obtener una respuesta, o un algoritmo de clasificación que permita obtener agrupaciones de información para así obtener hallazgos de

las bases de datos. Al igual que Big Data, requiere un conocimiento avanzado en programación, pero también conocimiento en modelos de aprendizaje.

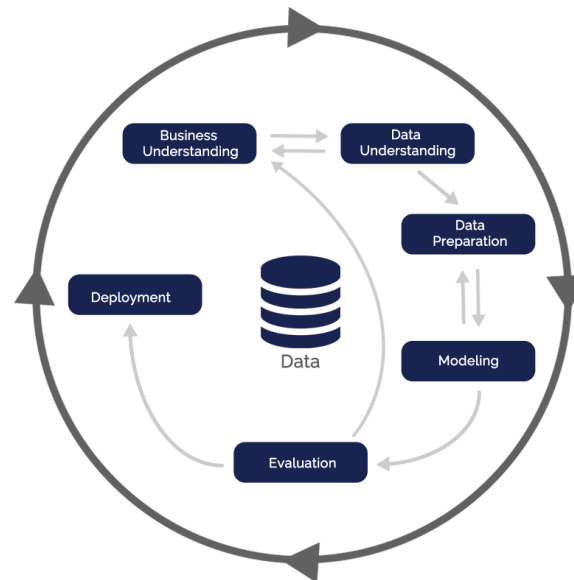
A continuación, se presenta una tabla comparativa de las opciones para poder abordar una solución a este problema:

Tabla 2. Tabla comparativa de opciones para abordar la solución

Método	Valor Económico Requerido	Conocimiento Teórico	Calidad De La Información	Tiempo Requerido	Calidad Del Resultado
Análisis De Mercado En Viñedos	Alto	Medio	Alta	Alto	Alta
Conocimiento De Expertos	Alto	Medio	Alta	Medio	Baja
Conocimiento De Agricultores	Medio	Bajo	Media	Medio	Baja
Big Data	Bajo	Medio	Alta	Medio	Alta
Inteligencia Artificial	Medio	Alto	Alta	Alta	Alta

Para este caso, se propone utilizar el análisis de Big Data, a través de la metodología CRISP-DM (*Cross Industry Standard Process for Data Miners*), debido a que es un proceso probado para el análisis de datos, que consiste en seis puntos principales los cuales se describen a continuación:

Figura 5. Metodología CRISP-DM



Nota. Adaptado de *What is The CRISP-DM Process?*, por Anirudh VK, 2019, AnalyticsIndiaMag.com (<https://analyticsindiamag.com/crisp-dm-data-science-project/>)

Entendimiento del Negocio (*Business Understanding*): Si duda, esta es la etapa más importante de toda la metodología. Es la etapa donde se debe entender el problema del negocio a resolver, cuáles son las variables gerenciales, económicas y administrativas impactadas, y que permita crear una estrategia, meta u objetivo conciso y específico (objetivo SMART, (Pastrana De La Cruz, 2019)).

Entendimiento de los Datos (*Data Understanding*): Una vez definida la meta, se debe validar si se tienen los correctos datos para poder obtener una respuesta propia para la misma. En esta etapa se debe validar los tipos de variables, si los datos que se tienen serán suficientes o se requieren otros adicionales de diferentes fuentes, si los datos son confiables, si están actualizados, explorar otro tipo de fuentes de datos o analizar y agregar nuevas variables que permitan fortalecer el análisis. En este punto es recomendable iterar con la etapa anterior, para así redefinir la meta u objetivos dependiendo de la disponibilidad de los datos.

Preparación de los datos (*Data Preparation*): Si las dos etapas anteriores han sido correctamente trabajadas y desarrolladas, el siguiente paso es el de obtener nuestros datos en un repositorio o carpeta de trabajo donde no puedan llegar a ser tratados, modificados o alterados. Se deben revisar las bases de datos para encontrar incongruencias, información incorrecta o incompleta. También se recomienda definir un tipo único formato para cada uno de los datos, en especial si hay valores de fechas, datos de medición bajo un sistema métrico definido y único (temperatura, peso, densidad, etc.), datos numéricos, etc.

Modelamiento (*Modeling*): Esta etapa describe el proceso de que se va a hacer con los datos donde cada etapa de este proceso tendrá una entrada y salida de datos, es decir, describe desde que se obtienen los datos, se preparan, procesan, conectan y se obtiene una salida (esto puede ser conclusiones estadísticas, modelos de análisis de ciencia de datos, segmentación de la información, etc.) En este punto es también importante iterar con la etapa anterior en la preparación de datos, ya que en algunos casos se terminan eliminando datos no relevantes o variables que no impactan en la meta del negocio. En esta etapa se seleccionan los modelos supervisados y no supervisados que permitan analizar los datos y generar conocimiento para la toma de decisiones relacionadas con el problema en estudio. Esta etapa también incluye la medición de la bondad del modelo, es decir, comparar la brecha entre los resultados con la realidad.

Evaluación (*Evaluation*): En esta etapa es que se evalúa si los datos modelados y su salida tienen sentido sobre la meta de negocio. Aquí es interesante entender que los resultados del modelamiento son el resultado de un proceso complejo, y que en esta etapa se debe analizar si dichos resultados podrán ser aplicados a la meta, por último, si están alineados a los valores, ética y objetivos de la empresa. Es por esto por lo que en esta etapa puede suceder que el planteamiento de la meta u objetivos del negocio puedan modificarse, reiniciando el proceso del modelo CRISP-DM.

Implementación (*Despliegue*): Esta última etapa permite usar la evaluación de datos para ponerla en práctica. Lo más interesante de esta es que nos ayuda a comprender falencias o problemas en el proceso general de la metodología, comprender si las variables

usadas fueron correctamente escogidas, preparadas, si el modelo y su evaluación fueron correctas. También demuestra si los resultados son acordes a la realidad o es necesario analizar otro tipo de variables que influyan en el mundo real.

Para poder aplicar la metodología CRISP-DM, se plantea realizar el siguiente proceso:

Primero, se plantea el problema, con la pregunta: ¿Cómo se puede determinar cuáles son los mejores rangos de las principales variables climáticas que influyen en la producción de un vino de alta calidad, para así poder tomar una decisión empresarial y de negocio para desplegar un nuevo viñedo?

La estrategia para resolver este problema será:

1. Estudiar las principales variables climáticas que afectan la calidad del vino
2. Obtener información global sobre estas variables
3. Estudiar que realmente es un vino de buena calidad
4. Obtener información global sobre vinos de buena calidad
5. Preparar, modelar y procesar en conjunto esta información
6. Evaluar estadísticamente la información obtenida
7. Realizar conclusiones sobre los resultados de la evaluación
8. Plantear opciones de como implementar las conclusiones obtenidas

5. Desarrollo del análisis de datos del estudio del caso

Siguiendo la metodología CRISP-DM, se describe cada paso y punto para el análisis completo de nuestro problema:

- a. Entendimiento del negocio.

En este apartado se cuestiona y analiza la pregunta principal de este proyecto con el análisis de los datos bajo la metodología CRISP-DM:

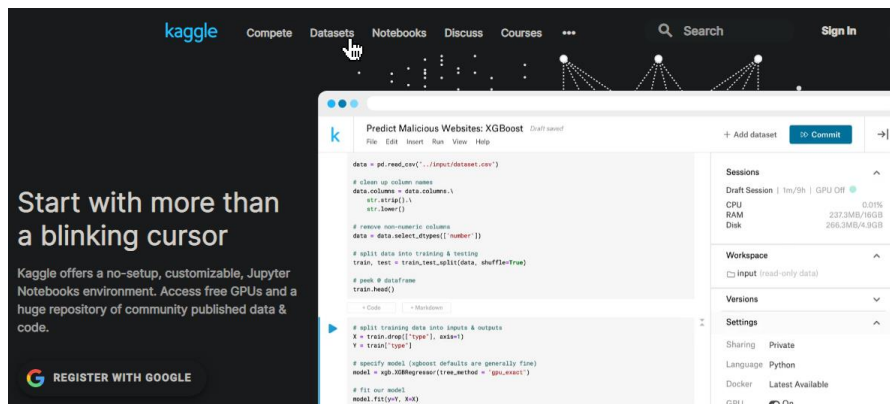
¿Cómo se puede determinar cuáles son los mejores rangos de las principales variables climáticas que influyen en la producción de un vino de alta calidad, para así poder tomar una decisión empresarial y de negocio para desplegar un nuevo viñedo? Partiendo de esto, la variable empresarial sería la de dar pautas para tomar la decisión de los rangos de las variables climáticas, y de una posible localización de un emprendimiento vinícola, aunque en el desarrollo de este análisis también puede llegar a incorporar otras variables dependiendo de las conclusiones, como decisión estratégica de la implementación de la cosecha, tecnologías a utilizar en la siembra, costos y materiales a usar, entre otros.

b. Entendimiento de los Datos

Para realizar un análisis de *Big Data*, requerimos de la información de origen para poder luego estudiarla. Entre las fuentes de información se pueden encontrar bases de datos o *datasets*, donde la primera es una colección organizada de datos que usualmente es gestionada a través de un sistema de gestión de bases de datos, y la segunda es básicamente información organizada en filas y columnas para ser procesada por *software* estadístico. Es posible que los *datasets* provengan de una base de datos, pero este no es un requerimiento.

El interés de este documento es del realizar el estudio de un *dataset* sobre calificaciones de vinos en diferentes regiones del mundo, con diferentes sepas y diferentes años. Existen múltiples fuentes gratuitas en Internet para obtener información tal como: el portal de datos abiertos de la Unión Europea (<https://open-data.europa.eu/es/data/>), el portal de datos abiertos del Gobierno de Estados Unidos (<https://www.data.gov/>), Fuentes de datos gratuitas de Amazon (<https://aws.amazon.com/datasets>), y muchos más. En este caso se usará *Kaggle*, una subsidiaria de Google, la cual es una comunidad online para científicos de datos, donde se pueden encontrar múltiples *datasets* de diversas áreas. En la página se pueden encontrar también módulos para análisis de *Big Data*, discusiones, cursos y blogs sobre *machine learning*, pero para el objetivo del documento solamente se centrará en *datasets*.

Figura 6. Página de inicio de kaggle.com



Nota. Adaptado de la página principal de Kaggle, por Kaggle, 2023, Kaggle.com (<https://www.kaggle.com/>)

Realizando la búsqueda con el término *wine*, se encuentran varios resultados, sin embargo, se utilizará el dataset con más usos y datos, el cual es *Wine Reviews* por *zackthoutt*, el cual obtuvo la información de *WineEnthusiast* (<http://www.winemag.com/>) en la semana del 22 de noviembre de 2017. En este caso se usará el *dataset winemag-data-130k-v2.csv* (alrededor de 130'000 entradas previamente filtradas para evitar duplicidad de los datos)

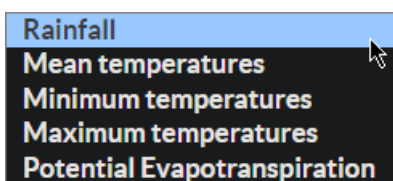
WineEnthusiast cuenta con alrededor de 300'000 entradas de datos y comentarios sobre vinos, el cual permite tener una gran cantidad de información al respecto. Su sede se encuentra en Valhalla, Nueva York y fue fundada en 1979, como una opción para dar recomendaciones de vinos. Esto permite entender de la fiabilidad de la información y del recorrido que tiene esta empresa, con respecto a los vinos y su calidad. La base de datos pura se puede encontrar en el siguiente enlace: <https://www.kaggle.com/zynicide/wine-reviews?select=winemag-data-130k-v2.csv>

Se propone, como tarea futura, utilizar otro *dataset* para encontrar correlación con los resultados de este análisis y así mismo, hacer un nuevo análisis de la misma información, pero actualizada al año presente.

Con esta única base de datos no es suficiente aun para poder realizar el análisis requerido. Para esto es requerida otra base de datos con información de latitud/longitud y variables climáticas, que permita unificar, evaluar y segmentar la información para así poder obtener datos concluyentes que permitan elaborar una respuesta certera a la pregunta principal del caso de este proyecto.

Para esto, se realizó una búsqueda similar de datos, donde se utiliza la información correspondiente a <https://www.globalclimatemonitor.org/>, el cual contiene información climática relacionada con la longitud y latitud a nivel global, con diferentes variables, las cuales se muestran a continuación:

Figura 7. Variables Climáticas por Tendencia y Longitud/Latitud de la fuente Global Climate Monitor



Nota. Adaptado de Global Climate Monitor, por la Universidad de Sevilla, 2023, GlobalClimateMonitor.org (<https://www.globalclimatemonitor.org/>)

Para el estudio se remueven datos atípicos los cuales no tenían información completa para poder ser usados en el análisis completo (por ejemplo, como vinos que no tenían información de su lugar de origen). Para el análisis se toman en cuenta las siguientes bases de datos con sus respectivos rangos:

- rainfall_anual_2000_2006.csv
- rainfall_anual_2007_2012.csv
- temp_anual_2000_2006.csv
- temp_anual_2007_2013.csv
- temp_maxima_2000_2006.csv
- temp_maxima_2007_2012.csv

- temp_minima_2000_2006.csv
- temp_minima_2007_2012.csv
- PET_anual_2000_2006.csv (Potential Evo Transpiration)
- PET_anual_2007_2012.csv (Potential Evo Transpiration)

Lamentablemente en las bases de datos trabajadas no se obtienen datos de más del 2013, sin embargo, en el tratamiento de datos se descartan los vinos cosechados en fechas por encima de 2012.

Para encontrar las bases de datos pura, se pueden descargar de la página <https://www.globalclimatemonitor.org/>, seleccionar la variable en el apartado “*By Trends*” y los respectivos años para descargar.

En resumen, se obtienen las siguientes variables para las respectivas bases de datos:

Base De Datos: *winemag-data-130k-v2.csv*

- **Id:** Identificador de la entrada en la base de datos, numérico
- **Country:** Nombre del país donde se produjo el vino, texto
- **Description:** Descripción del vino, como aroma, sabores, etc., texto
- **Designation:** Designación del jardín (*vineyard*) en el cual se cosechó el vino, texto
- **Points:** Puntaje de 1 a 100 sobre la calidad del vino, texto
- **Price:** Precio en dólares del costo de una botella de vino, numérico
- **Province:** Provincia o Estado donde se cosechó el vino, texto
- **Region_1:** Área de la provincia o del estado donde se cosechó el vino, texto
- **Region_2:** Área más específica donde se cosechó el vino, texto
- **Taster_name:** Nombre del catador quien dio la puntuación del vino, texto
- **Taster_twitter_handle:** Usuario de twitter del catador del vino, texto
- **Title:** Nombre del vino con el respectivo año de siembra, texto
- **Variety:** Varietal o uva correspondiente al vino, texto

- **Winery:** Nombre de la bodega vinícola donde se cosechó el vino, texto

Base De Datos: rainfall_anual_2000_2006.csv, rainfall_anual_2007_2012.csv

- **FID:** Dato referencial de la entrada en la base de datos, texto
- **Id_Punto:** Identificador del punto de localización, numérico
- **Agno:** Año del dato de la precipitación Anual, numérico
- **Pre_anual:** Precipitación anual promedio, numérico en mm
- **Geom:** Datos en valor WGS84, texto

Base de Datos: temp_anual_2000_2006.csv, temp_anual_2007_2013.csv

- **FID:** Dato referencial de la entrada en la base de datos, texto
- **Id_Punto:** Identificador del punto de localización, numérico
- **Agno:** Año del dato de la precipitación Anual, numérico
- **temp_anual:** Temperatura anual promedio, numérico en centígrados
- **Geom:** Datos en valor WGS84, texto

Base de Datos: temp_maxima_2000_2006.csv, temp_maxima_2007_2012.csv

- **FID:** Dato referencial de la entrada en la base de datos, texto
- **Id_Punto:** Identificador del punto de localización, numérico
- **Agno:** Año del dato de la precipitación Anual, numérico
- **temp_max_anual:** Temperatura anual máxima, numérico en centígrados
- **Geom:** Datos en valor WGS84, texto

Base de Datos: temp_minima_2000_2006.csv, temp_minima_2007_2012.csv

- **FID:** Dato referencial de la entrada en la base de datos, texto
- **Id_Punto:** Identificador del punto de localización, numérico
- **Agno:** Año del dato de la precipitación Anual, numérico
- **temp_min_anual:** Temperatura anual mínima, numérico en centígrados
- **Geom:** Datos en valor WGS84, texto

Base de Datos: PET_anual_2000_2006.csv, PET_anual_2007_2012.csv

- **FID:** Dato referencial de la entrada en la base de datos, texto
- **Id_Punto:** Identificador del punto de localización, numérico

- **Agno:** Año del dato de la precipitación Anual, numérico
- **etp_anual:** Evo transpiración, numérico en milímetros
- **Geom:** Datos en valor WGS84, texto

Estas variables son suficientes para el análisis que tiene como objetivo este estudio, ya que la meta es poder obtener la relación de puntuaciones de vino con su provincia, región, o estado para así conectarlo con las respectivas coordenadas geográficas, para así obtener la información de las variables climáticas relacionadas con el año de producción del vino.

c. Preparación de los Datos

En los anexos a este documento se presentan los respectivos análisis y procesamiento de los datos para obtener una única base de datos a trabajar. Una descripción general de esta preparación de datos se encuentra en el documento “*Summary Preproc y EDA.docx*”

En resumen, el proceso para la preparación de datos se realizaron los siguientes pasos:

Base de Datos sobre calificación de vinos

1. De la base de datos principal de la calificación de vino se conservaron los datos de puntaje, precio, año de cosecha del vino, país y región.
2. Se eliminaron valores atípicos, que no presentaban datos correctos o que no incluían en su información datos de las variables requeridas en el punto anterior
3. Se realiza una unión de datos del país, provincia, región_1 y región_2, para poder establecer lo más precisamente posible la ubicación donde se cosechó el vino.
4. Se usa GoogleV3 API como herramienta de conversión de datos de la ubicación en texto a localización de Latitud y Longitud.

Base de Datos sobre variables climáticas

1. En todas las bases de datos de información climática se hace realiza la conversión de datos “*Geom*”, que están en formato WGS84 a las Columnas Latitud y Longitud.

Posterior a este procesamiento de datos se realiza unión de todas las bases de datos. Las variables para unir estas bases de datos son:

Año de cosecha del vino, el cual se obtiene de la variable “*Title*”, la cual se conecta con el valor del año en cada una de las bases de datos climatológicos.

Latitud y Longitud (*país, provincia, región_1, región_2*), la cual se conecta con las variables Latitud y Longitud de las bases de datos climatológicos.

Esto resulta en una base de datos única con un total de 70’030 entradas, la cual es llamada “*Dataframe_final*”, como se puede apreciar en la siguiente figura

Figura 8. Resumen de base de datos final con variables principales.

	points	price	Year	Lat_x	Long_x	temp_anual	temp_max_anual	temp_min_anual	pre_anual	etp_anual	country_Arg
0	87	15.0	2011	41.75	-5.75	13.01	19.44	6.61	388.5	1200	0
1	87	15.0	2011	41.75	-5.75	13.01	19.44	6.61	388.5	1200	0
2	87	17.0	2011	41.75	-5.75	13.01	19.44	6.61	388.5	1200	0
3	91	12.0	2011	41.75	-5.75	13.01	19.44	6.61	388.5	1200	0
4	87	8.0	2011	41.75	-5.75	13.01	19.44	6.61	388.5	1200	0
...
70025	84	25.0	2012	34.75	-118.25	13.96	20.23	7.75	325.0	1176	0
70026	90	21.0	2012	46.75	6.75	9.15	13.46	4.87	1382.5	696	0
70027	89	14.0	2012	45.25	6.25	5.61	9.20	2.05	1386.2	714	0
70028	89	18.0	2012	45.75	5.75	10.71	15.25	6.22	976.2	807	0
70029	87	25.0	2012	40.25	15.25	15.09	18.80	11.40	269.6	1077	0

70030 rows × 12 columns

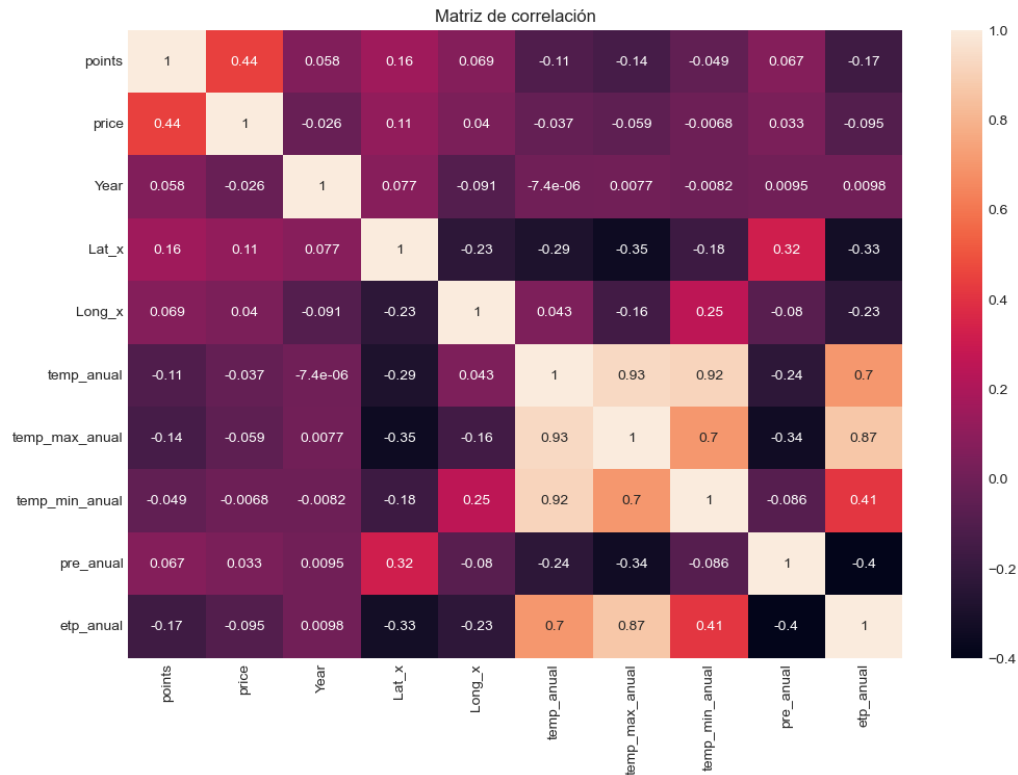
Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

d. Modelamiento

En el documento anexo “*EDA_winereviews_RGSB.pdf*” se encuentra el análisis estadístico con las respectivas conclusiones. Se recomienda ver el documento para entender el resumen a continuación:

En esta matriz de correlación se puede ver en grandes rasgos, cuales variables tienen una correlación directa (positiva, entre más cercano a 1, mayor correlación) o correlación inversa (negativa, entre más cercano a -1, menor correlación).

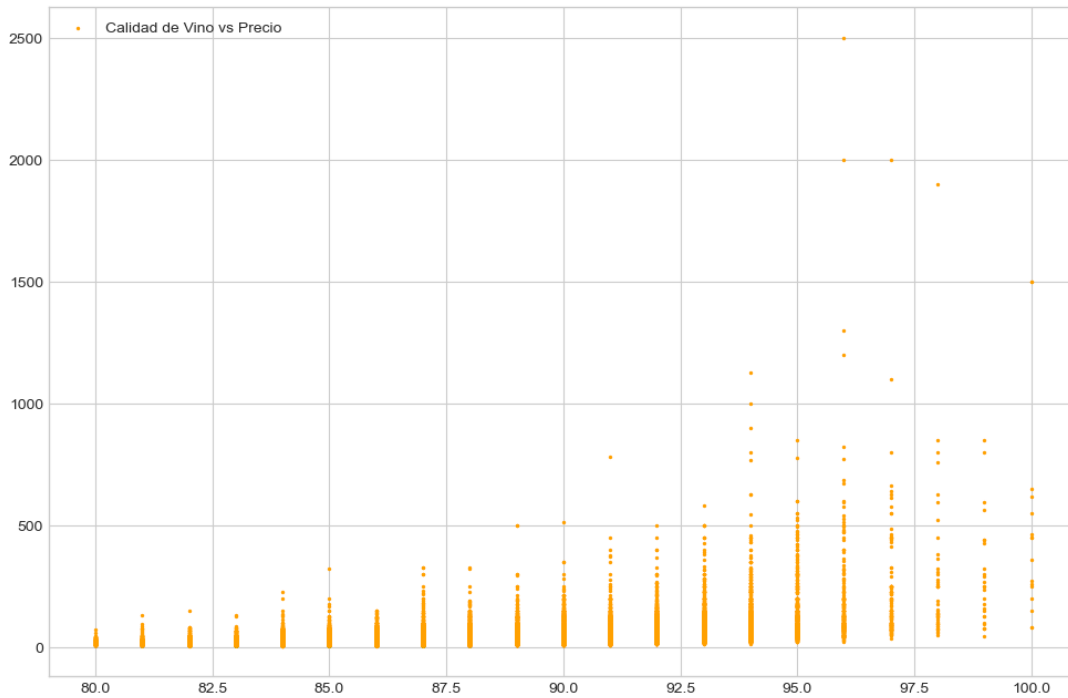
Figura 9. Matriz de Correlación



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

Inicialmente se encuentra que hay una correlación media entre el puntaje de los vinos y su precio. Este análisis se hace sobre toda la base de datos por lo que esta información no es concluyente para nuestro objetivo final.

Figura 10. Relación de Calidad de vino (1-100) vs Precios (USD por botella)



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

Aunque con el análisis inicial de la base de datos se obtienen algunas conclusiones, se procede a realizar un análisis ANOVA, el cual permite entender la varianza de los datos para así determinar si existe alguna diferencia muy alta entre las “medias” de los diferentes grupos.

Por último, se puede ingresar a estos dos repositorios de *GitHub*, donde se puede encontrar el análisis completo del desarrollo procedimental en *Jupyter*.

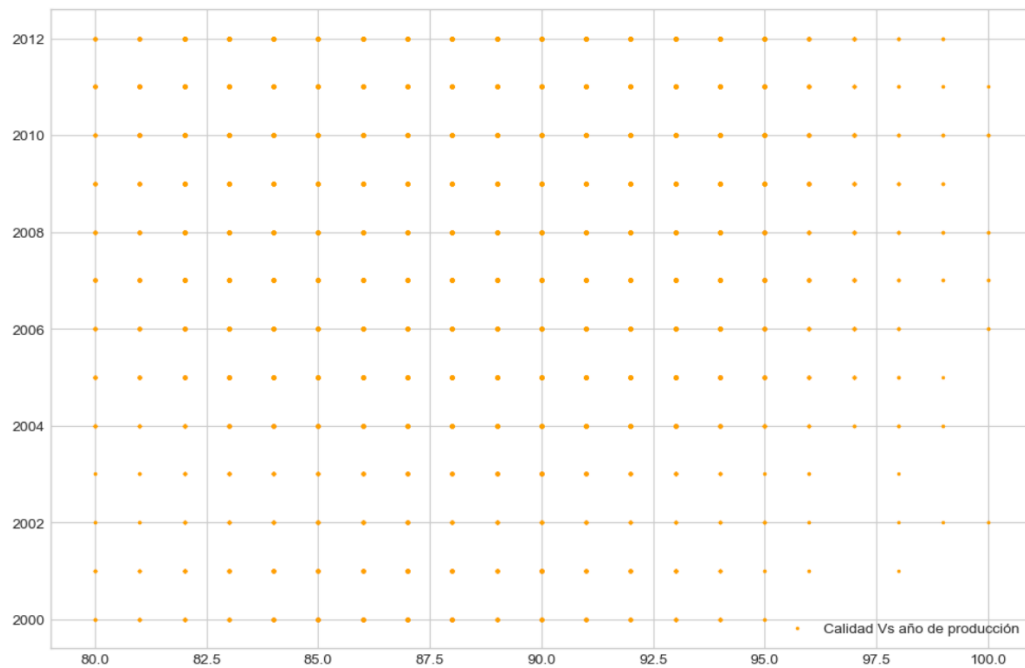
- <https://github.com/crcoronado/winetesis>
- <https://github.com/crcoronado/Datascience1>

e. Evaluación

En esta evaluación de los datos obtenidos, se tienen las siguientes conclusiones:

- Existe una relación entre la calidad del vino y el precio
- No existe una relación entre el año de cosecha del vino y la calidad, sin embargo, resaltan algunos años como 2002, 2006, 2007, 2008, 2010 y 2011 en vinos de calidad 100.

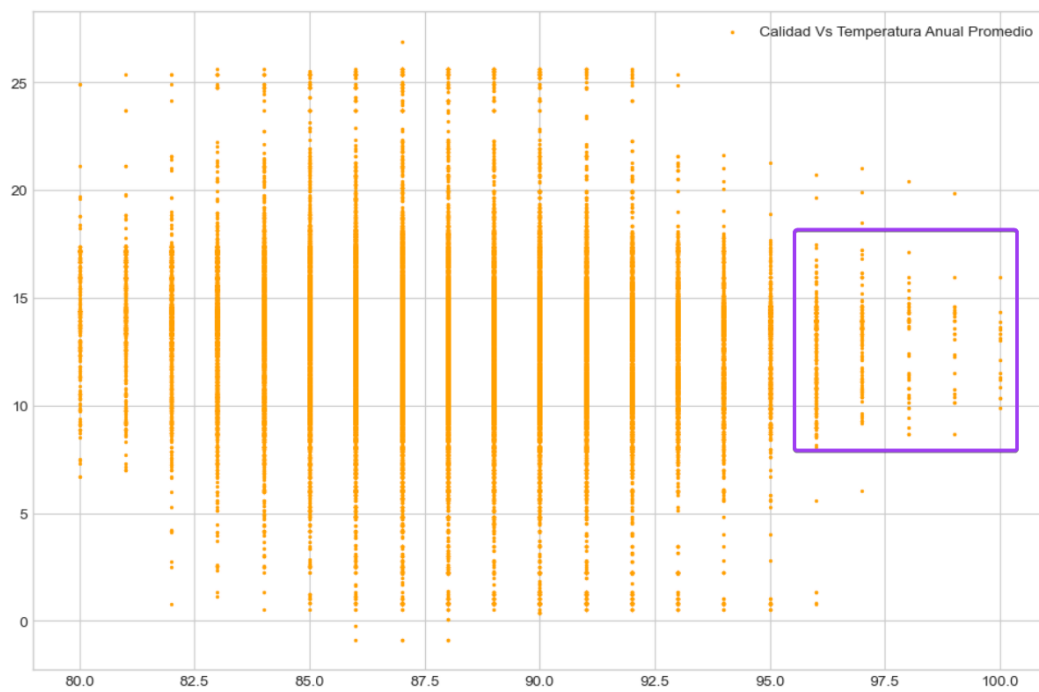
Figura 11. Relación entre año de producción y calidad del vino



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Existe una segmentación en que los mejores vinos son cosechados en temperaturas promedio al año entre los 9°C y 16 °C, concentrada especialmente entre los 10°C y 14,5 °C.

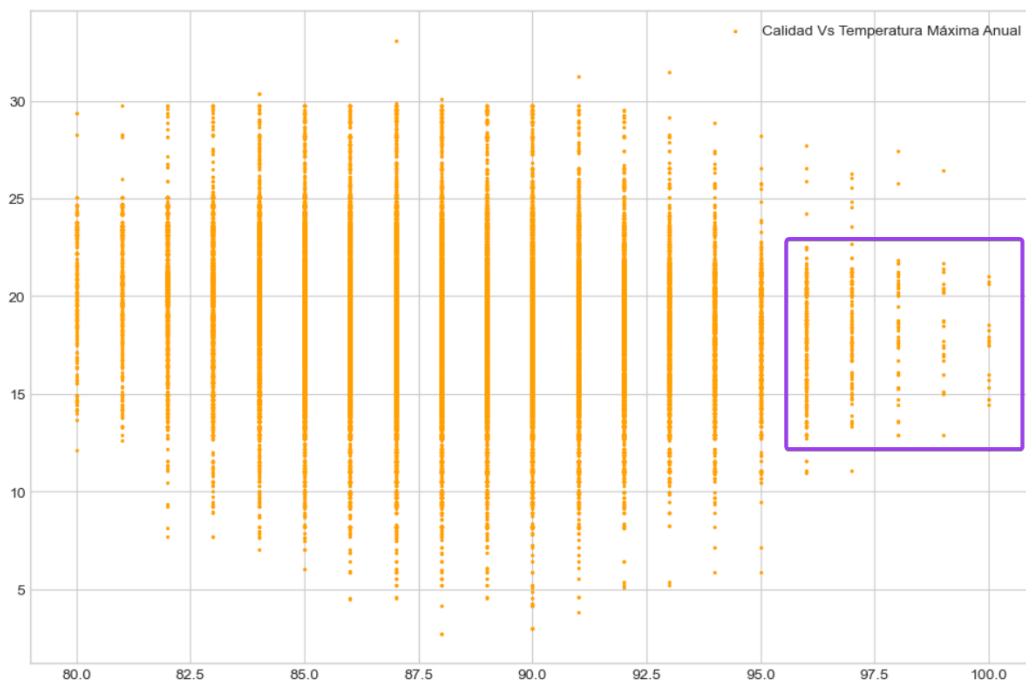
Figura 12. Relación entre Calidad del Vino y Temperatura anual promedio de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Existe una segmentación en que los mejores vinos son cosechados en temperaturas máximas entre los 14 °C y 22 °C.

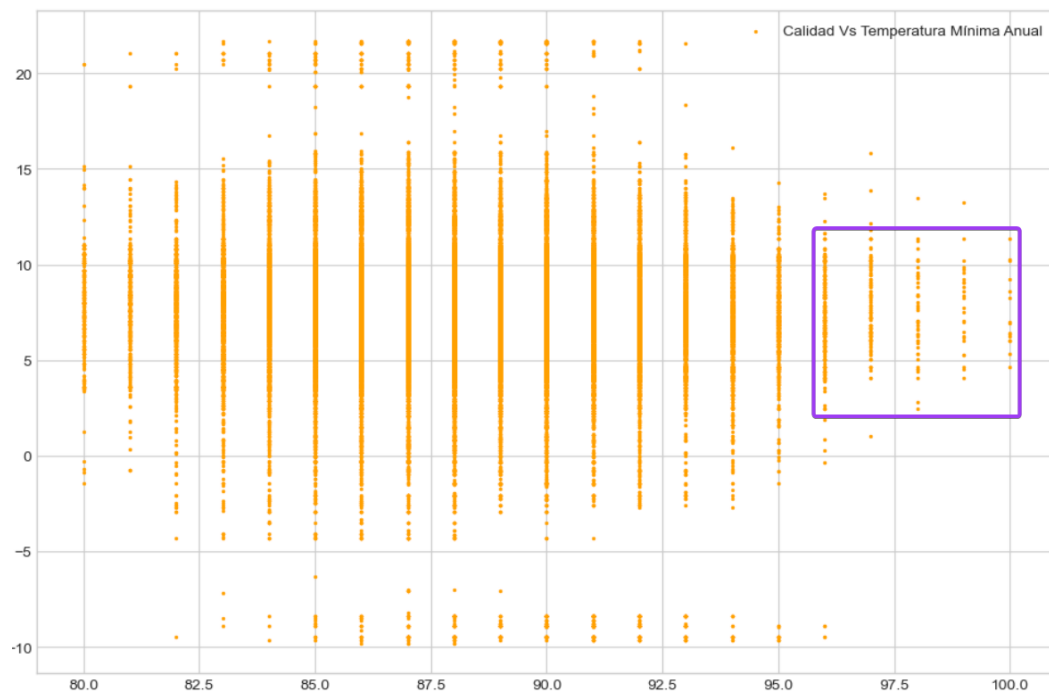
Figura 13. Relación entre la Calidad del Vino y Temperatura anual máxima de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Existe una segmentación en que los mejores vinos son cosechados en temperaturas mínimas entre los 11 °C y 4 °C

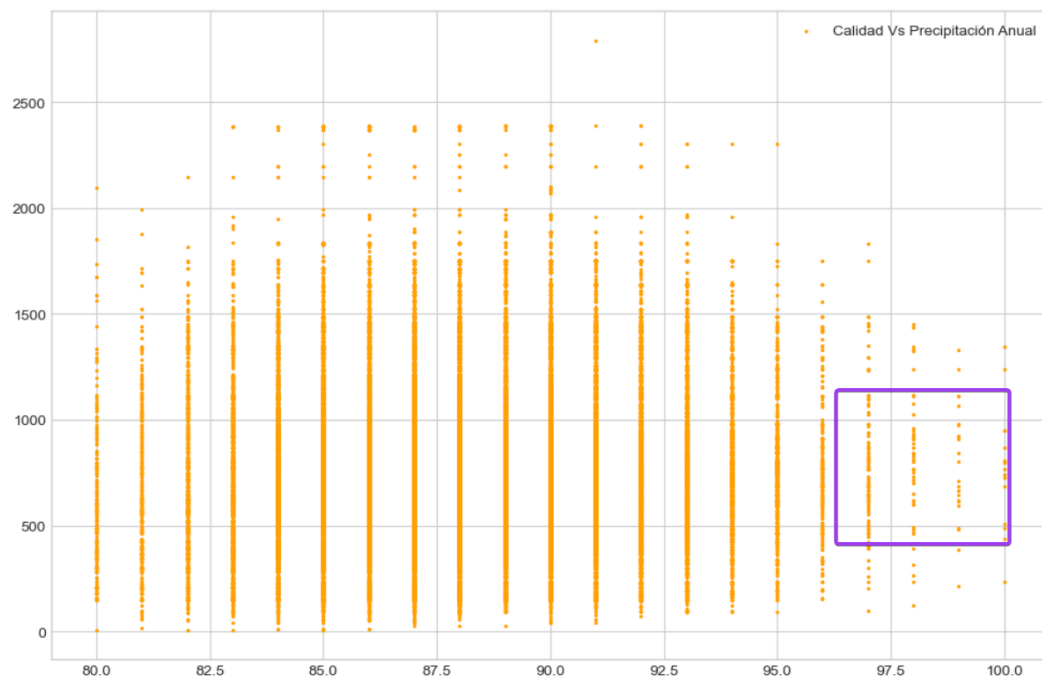
Figura 14. Relación entre la Calidad del Vino y la Temperatura Mínima Anual de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Existe una segmentación que los mejores vinos son cosechados en lugares con precipitación promedio entre los 450mm y los 1100mm al año

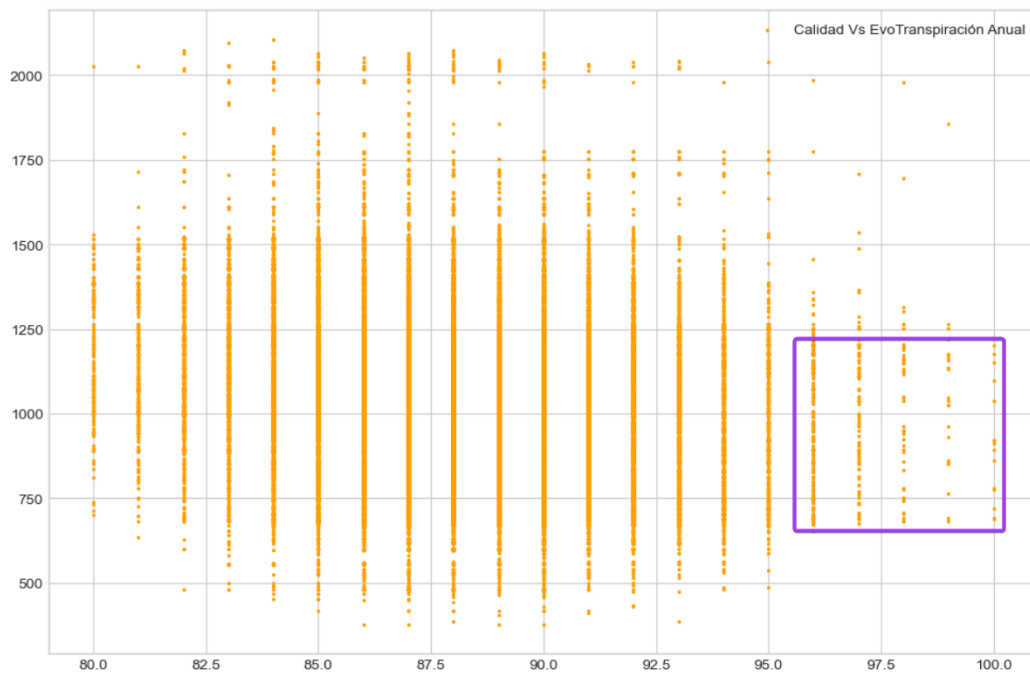
Figura 15. Relación entre Calidad del Vino y Precipitación Anual de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Existe una segmentación de que los mejores vinos son cosechados en lugares con Evapotranspiración promedio entre los 750mm y 1250mm al año

Figura 16. Relación entre Calidad del Vino y Evo Transpiración Anual promedio de la zona de la cosecha

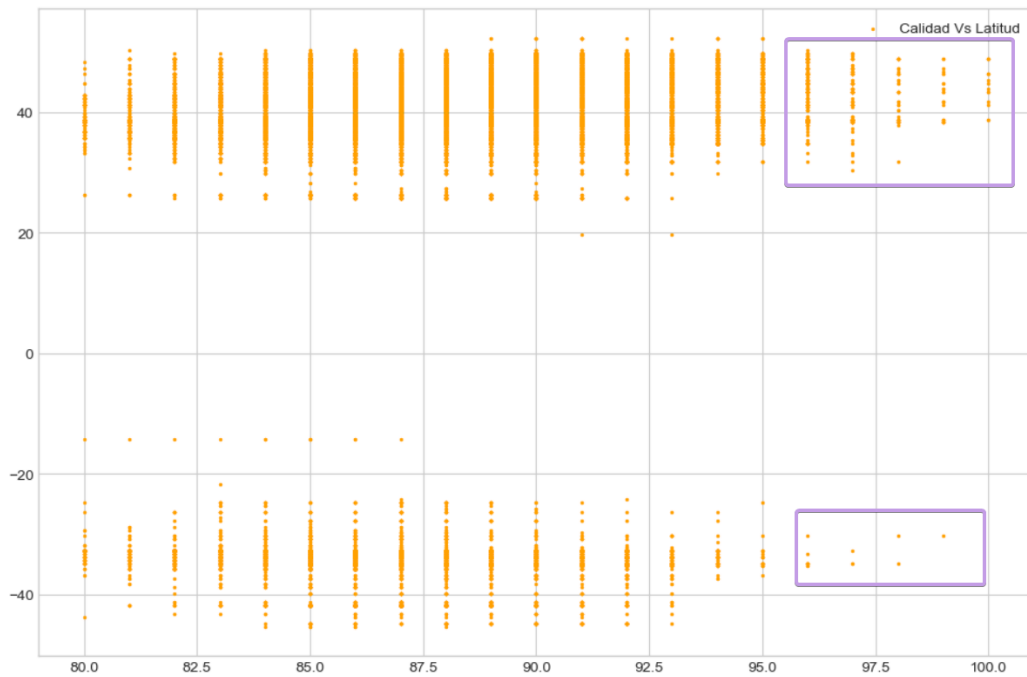


Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

Algunas otras conclusiones del análisis que se encuentra son:

- La mayoría de los vinos de buena calidad se encuentran sobre el trópico de Cáncer y algunos en el trópico de Capricornio, sin encontrar valores en la zona de los trópicos de Ecuador. Esto principalmente por que la uva necesita de un rango de temperatura para poder producir un vino de buena calidad.

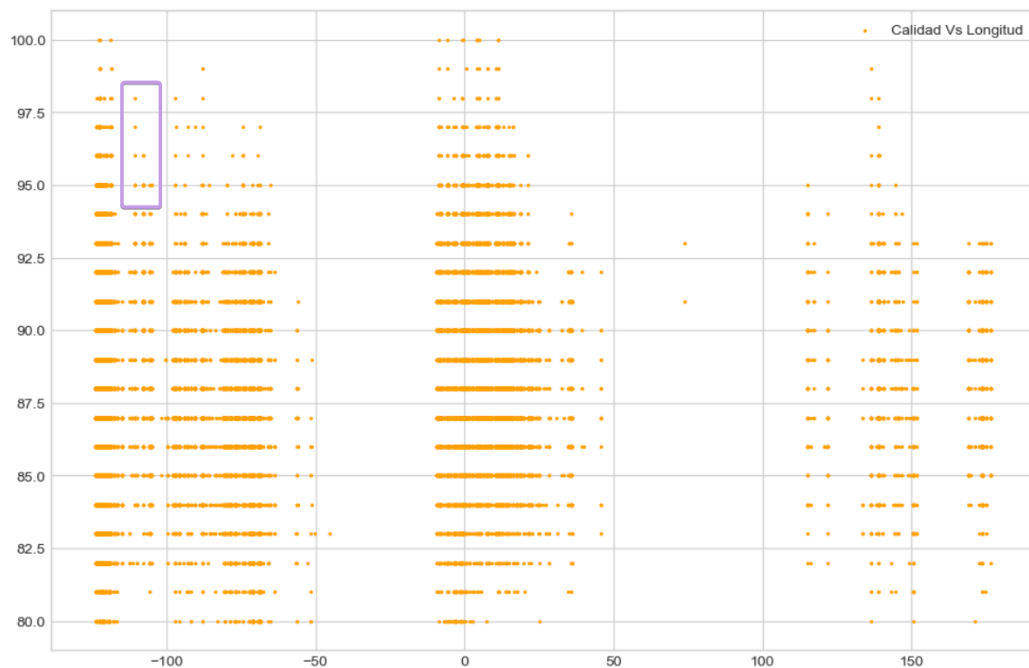
Figura 17. Relación entre Calidad del Vino y Latitud de la región de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

- Los mejores vinos se encuentran cosechados en la región de California, sin embargo, se encontraron algunos en la zona de Longitud -111° a -107° . Esto permitiría analizar las zonas más interesantes para poder implementar una zona de cosecha de la vid.

Figura 18. Relación entre Calidad del Vino y Longitud de la zona de la cosecha



Nota. Adaptado del análisis realizado en Jupyter, la cual se puede encontrar en el archivo Anexo.

Regresando a la pregunta principal: “¿Cómo se puede determinar cuáles son los mejores rangos de las principales variables climáticas que influyen en la producción de un vino de alta calidad, para así poder tomar una decisión empresarial y de negocio para desplegar un nuevo viñedo?”

La respuesta es que a través de la ciencia de datos y basándonos en análisis de *Big Data* se encontraron una segmentación de rangos en relación con los vinos de mejor calidad, la cual permite plantear nuevas zonas como opciones de viñedos para nuevos negocios.

6. Plan y recomendaciones de implementación y aplicación

f. Implementación

Siguiendo el modelo CRIP-DM, el alcance de la implementación de estas conclusiones se proponen para una nueva fase de desarrollo.

Uno de los puntos propuestos es el de enviar esta información al viñedo donde se realizó la visita de negocio en Barcelona, el cual fue acompañado por el grupo del EADA, con los siguientes pasos:

- Monitorear las variables climáticas sobre las cosechas de vinos, segmentando cuales se encuentran dentro o cerca de los rangos de las conclusiones del estudio
- Analizar si las cosechas que se acercan a los rangos presentan una calidad mayor que la de otras cosechas
- Controlar, en la cosecha, una porción del viñedo, controlando las variables climáticas, para así, poder producir dos vinos finales para poder, a través de *sommeliers* o catadores, si se presenta una diferencia notable en la calidad del vino

7. Hallazgos

Sobre este proyecto, se presentan los siguientes hallazgos.

- Las segmentaciones de los rangos obtenidos permiten que los viñedos mejoren sus prácticas y así mantener los cultivos entre los rangos presentados, lo que permitiría validar en la práctica las conclusiones estadísticas del *Big Data*.
- La obtención de información y datos a través de un análisis de grandes cantidades de datos impulsa a los agricultores y cosechadores de vino a mejorar sus prácticas, calidad del producto final y ganancias a través del tiempo.
- Incentiva a los cosechadores y agricultores a utilizar nuevas tecnologías y métodos de monitoreo, que permita no solo obtener la información en tiempo real, sino también guardarla y procesarla para obtener nuevos hallazgos.
- Permite a la agricultura vinícola, implementar procesos de control en caso de que los pronósticos climatológicos indiquen alguna anomalía o situación particular que afecte el desarrollo del fruto.
- Presenta a emprendedores o viñedos ya establecidos, nuevas opciones de zonas que cuenten con rangos cercanos a los mostrados para poder establecer planes piloto y pruebas, aplicando un enfoque diferente al ya conocido.
- Expone nuevas opciones de cultivo de la vid, tal como buscar zonas de tierra propicia para el cultivo de uvas, y con la ayuda de nuevas tecnologías que permiten el control de temperatura, humedad e irrigación, implementar cultivos para la obtención de un fruto de alta calidad.
- Así mismo, el estudio técnico presenta el potencial de la ciencia de datos para poder inferir y/o concluir información estratégica, relacionada con la toma de decisiones de negocio, procesos, empresariales y gerenciales.
- La metodología CRISP-DM usada en la ciencia de datos para análisis de información es muy similar a la metodología basada en retos. Esto permite que los procesos, análisis y cuestionamientos sean similares en este proyecto

- Presenta a la comunidad del RGSB que es imperativo que se dé una introducción a la ciencia de datos en el *Master of Business Administration*, ya que es una herramienta que se utiliza con mayor frecuencia en la toma de decisiones empresariales.
- Usar el método de aprendizaje basado en retos permite desarrollar y seguir con fluidez un proceso de análisis para así poder llegar a un desarrollo, modelo, implementación y conclusiones objetivas.
- Finalmente, expone que las clases de la maestría permiten construir bases y reforzar conocimientos para preparar a los estudiantes a asumir retos administrativos y gerenciales, liderar equipos de dirección y administración, y actualizarse en nuevas tendencias de conocimiento.

8. Referencias bibliográficas

- 10 Most Innovative Data Science Companies in 2023.* (s/f). Recuperado el 19 de agosto de 2023, de <https://www.knowledgehut.com/blog/data-science/data-science-companies>
- 43 Best Data Science MBA Programs & Data Analytics MBA.* (s/f). DiscoverDataScience.Org. Recuperado el 19 de agosto de 2023, de <https://www.discoverdatascience.org/programs/mba-data-science/>
- Bai, H., Gambetta, G. A., Wang, Y., Kong, J., Long, Q., Fan, P., Duan, W., Liang, Z., & Dai, Z. (2022). Historical long-term cultivar×climate suitability data to inform viticultural adaptation to climate change. *Scientific Data*, 9(1), Article 1. <https://doi.org/10.1038/s41597-022-01367-6>
- Big Data and Agriculture: A Complete Guide.* (s/f). Talend - A Leader in Data Integration & Data Integrity. Recuperado el 19 de agosto de 2023, de <https://www.talend.com/resources/big-data-agriculture/>
- Corvain. (2022, marzo 25). *Southern Hemisphere Wines: Facts and Wines from Coravin / Coravin.* <https://www.coravin.com/community/southern-hemisphere-wines-facts-and-wines-from-coravin>
- Crisp DM methodology. (s/f). *Smart Vision Europe.* Recuperado el 19 de agosto de 2023, de <https://www.sv-europe.com/crisp-dm-methodology/>
- Dutta, A. (2004). *Big Data Analytics Changing the Centuries-Old Wine Industry.* Sommeliers Choice Awards. <https://sommelierschoiceawards.com/en/blog/insights-1/big-data-analytics-changing-the-centuries-old-wine-industry-684.htm>

- Echeverri, L. (2019, junio 24). *Part Time—Presentación Dirección de Marketing 2019*.
- Frick, W. (2018, enero 22). 3 Ways to Improve Your Decision Making. *Harvard Business Review*. <https://hbr.org/2018/01/3-ways-to-improve-your-decision-making>
- Gambetta, G. A., & Kurtural, S. K. (2021). Global warming and wine quality: Are we close to the tipping point? *OENO One*, 55(3), Article 3. <https://doi.org/10.20870/oeno-one.2021.55.3.4774>
- Globe Newswire. (2020, septiembre 8). *Global Wine Industry*. Benzinga; Benzinga. <https://www.benzinga.com/pressreleases/20/09/g17403336/global-wine-industry>
- Hayashi, A. M. (2001, febrero). *When to Trust Your Gut*. When-to-Trust-Your-Gut. <https://hbr.org/2001/02/when-to-trust-your-gut>
- Hillier, W. (2021, septiembre 20). *4 Fascinating Real-World Big Data Examples*. <https://careerfoundry.com/en/blog/data-analytics/big-data-examples/>
- IntelligentHQ. (2020, agosto 20). Why And How is Big Data Impacting The Hospitality Industry? *IntelligentHQ*. <https://www.intelligenthq.com/big-data-impacting-hospitality-industry/>
- Jackson, R. S. (2017). Chapter 8—Nature and Origins of Wine Quality. En R. S. Jackson (Ed.), *Wine Tasting (Third Edition)* (pp. 337–370). Academic Press. <https://doi.org/10.1016/B978-0-12-801813-2.00008-2>
- Jay, A. (2019, septiembre 11). *97 Big Data Statistics You Must Learn: 2023 Market Share & Data Analysis*. Financesonline.Com. <https://financesonline.com/big-data-statistics/>

- Jones, G., White, M., Cooper, O., & Storchmann, K. (2005). Climate Change and Global Wine Quality. *Clim Change*, 73, 319–343. <https://doi.org/10.1007/s10584-005-4704-2>
- La influencia de la luna en la agricultura—Colombia Verde*. (2023, mayo 11). <https://colombiaverde.com.co/geografia/agricultura/la-influencia-de-la-luna-en-la-agricultura/>
- Leon, J. (2017, mayo 11). El ciclo de la vid ilustrado paso a paso. *Jean Leon*. <https://www.jeanleon.com/el-ciclo-de-la-vid-ilustrado-paso-a-paso/>
- Marland, J. (2014, abril 30). *SAP BrandVoice: How Big Data Can Predict The Wine Of The Century*. Forbes. <https://www.forbes.com/sites/sap/2014/04/30/how-big-data-can-predict-the-wine-of-the-century/>
- McKay, A. (Director). (2015, diciembre 23). *The Big Short* [Biography, Comedy, Drama, History]. Paramount Pictures, New Regency Productions, Plan B Entertainment.
- Nguyen, D. (2020, noviembre 25). *Red Wine Quality Prediction Using Regression Modeling and Machine Learning*. Medium. <https://towardsdatascience.com/red-wine-quality-prediction-using-regression-modeling-and-machine-learning-7a3e2c3e1f46>
- Pandey, K. (2022, julio 14). *5 Real-World Examples of Data Analytics*. Masai School. <https://www.masaischool.com/blog/5-real-world-examples-of-data-analytics/>
- Pastrana De La Cruz, M. (2019, febrero 21). *2a Part Time—Presentación Administración 2019*.
- Predictive Analytics: Grow your business like Netflix!* (s/f). Engati. Recuperado el 19 de agosto de 2023, de <https://www.engati.com/blog/predictive-analytics>

- Python or R for Data Analysis: Which Should I Learn?* (2023, junio 15). Coursera.
<https://www.coursera.org/articles/python-or-r-for-data-analysis>
- ¿Qué es la ciencia de datos? - Explicación de la ciencia de datos - AWS.* (s/f). Amazon Web Services, Inc. Recuperado el 30 de julio de 2023, de <https://aws.amazon.com/es/what-is/data-science/>
- Santos, D. (2023, enero 20). *Qué es un análisis de mercado, cómo se hace y ejemplos.*
<https://blog.hubspot.es/marketing/como-hacer-analisis-mercado>
- The Deciding Factor: Big Data & Decision Making. (2012, junio 4). *Capgemini Worldwide.*
<https://www.capgemini.com/resources/the-deciding-factor-big-data-decision-making/>
- Tudor, N. (2021, febrero 10). *7 real-world examples of how brands are using Big Data analytics.* Bornfight. <https://www.bornfight.com/blog/7-real-world-examples-of-how-brands-are-using-big-data-analytics/>
- Vaughan, A. (2023, junio 23). *Top Big Data Tools & Software for 2023.* TechnologyAdvice.
<https://technologyadvice.com/blog/information-technology/big-data-tools/>
- Vinos y Sabores.* (s/f). Recuperado el 20 de agosto de 2023, de <http://www.vinosygastronomia.com/variedades>
- Webb, A. (2018, octubre 26). *12 Wines From Countries You Didn't Even Know...* Culture Trip. <https://theculturetrip.com/europe/united-kingdom/articles/12-wines-from-countries-you-didnt-even-know-made-wine/>
- What Is Big Data?* (s/f). Google Cloud. Recuperado el 21 de agosto de 2023, de <https://cloud.google.com/learn/what-is-big-data>

What the Wine Industry Understands About Connecting with Consumers. (2019, marzo 5).

Harvard Business Review. <https://hbr.org/2019/03/what-the-u-s-wine-industry-understands-about-connecting-with-customers>

Where Did Wine Come From? The True Origin of Wine. (s/f). Wine Folly. Recuperado el 4 de julio de 2022, de <https://winefolly.com/deep-dive/where-did-wine-come-from/>

Wine (HS: 2204) Product Trade, Exporters and Importers. (2018). <https://oec.world/en/profile/hs92/wine>

Yanet Acosta et al. (2013). *El mundo del vino*. LAROUSEE EDITORIAL.