



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey

Escuela de Ingeniería y Ciencias

Maestría en Inteligencia Artificial Aplicada

Dra. Grettel Barceló Alonso

Dr. Luis Eduardo Falcón Morales

Materia Proyecto integrador

Análisis exploratorio de datos

Equipo 11

José Gilberto Amaro Aceves - A01794815

Francisco Antonio Enríquez Cabrera - A01795006

Carlos Eduardo García García - A01795206

04-05-2025

Índice

Liga Google Colab	4
Repositorio GitHub	4
Introducción	4
Exploración de datos	4
¿Como fueron obtenidos los datos?	4
¿Propiedades encontradas para analizar?	5
a. Consumo de agave y producción totales del tequila	6
b. Exportaciones por país	7
c. Exportaciones Total de Forma	11
d. Total de Exportaciones Categoría	12
Preprocesamiento	13
Corrección de nulos	13
Reducción de cardinalidad	14
Borrado de datos innecesarios	14
Descubrimientos	15
Valores faltantes / ausencia	15
Correlaciones	17
Valores atípicos	19
Tendencias temporales	22
a. Exportaciones Total de Forma	22
b. Total de Exportaciones Categoría	23
c. Producción Total	24
d. Consumo de Agave Total	25
e. Exportaciones por País	29

Desequilibrios	30
Conclusión	32

Liga Google Colab

https://colab.research.google.com/drive/1Mswc86ZaltoDcGeR6p-8dvFIXLlkZhSl?usp=chrome_ntp

Repositorio GitHub

Repositorio: <https://github.com/FranciscoEnriquez/Proyecto-Modelo-de-Pronostico-de-Exportaciones>

Notebook con EDA: <https://github.com/FranciscoEnriquez/Proyecto-Modelo-de-Pronostico-de-Exportaciones/blob/main/notebook/eda.ipynb>

Introducción

En este proyecto se tiene como objetivo analizar los datos públicos de la Cámara Nacional de la Industria Tequilera (CNIT) para obtener un modelo de proyección de exportaciones y un segundo modelo para la proyección del consumo de agave.

En este avance se abordará todo lo relacionado a nuestras actividades de exploración de datos, tal como el diccionario, estadísticas de nuestros datos, frecuencia en nuestras variables categóricas, correlaciones, etc.

También se abarcará el preprocesamiento de nuestros datos, los descubrimientos obtenidos durante esta exploración y se presentarán las conclusiones.

Exploración de datos

¿Como fueron obtenidos los datos?

Los datos fueron obtenidos mediante los reportes públicos de la misma Cámara Nacional de la Industria Tequilera en el siguiente enlace: <https://old.crt.org.mx/EstadisticasCRTweb/>

Se hizo un programa el cual ejecuta un web scrapping de estos reportes para obtener los conjuntos de datos necesarios para nuestro trabajo, dichos conjuntos de datos siendo los siguientes:

1. Consumo de agave total
2. Exportaciones por país
3. Exportaciones totales por categoría
4. Exportaciones totales por forma
5. Producción total de tequila.

¿Propiedades encontradas para analizar?

En esta sección se abodarán los datos disponibles en cada uno de los conjuntos de datos, sin embargo, todos los conjuntos de datos tienen en común con que son “series de tiempo” que el negocio va a actualizando su registro el día 8 de cada mes.

En general la variable objetivo siempre será el Valor que se intentará predecir en entregas futuras.

a. Consumo de agave y producción totales del tequila

En este conjunto de datos se cuenta con una matriz de 724 registros/filas y 5 columnas, de las cuales, al

Columna	Tipo de Dato	Clasificación	Descripción
SubCategoria	object	Catagórica nominal	Categoría del tequila usado para el consumo del agave (ej. Tequila 100%, Tequila).
Year	int64	Numérica continua	Año del registro del consumo de agave
Valor	float64	Numérica	Consumo de agave en miles de toneladas.
AñoArchivo	int64	Numérica continua	Año del archivado del registro del consumo de agave
Mes	int64	Numérica continua	Mes donde fue consumido

Tabla 1. Diccionario de datos de consumo de agave.

De estos conjuntos de datos se obtuvieron los siguientes resúmenes descriptivos:

Resumen descriptivo:						
	SubCategoria	Year	Valor	AnioArchivo	Mes	
count	724	724.000000	724.000000	724.000000	724.000000	
unique	2	NaN	NaN	NaN	NaN	
top	Tequila	NaN	NaN	NaN	NaN	
freq	362	NaN	NaN	NaN	NaN	
mean	NaN	2009.585635	11.392127	2009.585635	6.472376	
min	NaN	1995.000000	0.700000	1995.000000	1.000000	
25%	NaN	2002.000000	7.275000	2002.000000	3.000000	
50%	NaN	2010.000000	10.300000	2010.000000	6.000000	
75%	NaN	2017.000000	13.325000	2017.000000	9.000000	
max	NaN	2025.000000	47.800000	2025.000000	12.000000	
std	NaN	8.713647	7.594253	8.713647	3.464989	

Figura 1. Resumen descriptivo, producción total de tequila.

```

Resumen descriptivo:
      SubCategoria      Year      Valor      AñoArchivo      Mes
count      724      724.000000      724.000000      724.000000      724.000000
unique      2      NaN      NaN      NaN      NaN
top      Tequila      NaN      NaN      NaN      NaN
freq      362      NaN      NaN      NaN      NaN
mean      NaN      2009.585635      41.197790      2009.585635      6.472376
min      NaN      1995.000000      3.700000      1995.000000      1.000000
25%      NaN      2002.000000      21.400000      2002.000000      3.000000
50%      NaN      2010.000000      28.500000      2010.000000      6.000000
75%      NaN      2017.000000      45.225000      2017.000000      9.000000
max      NaN      2025.000000      219.300000      2025.000000      12.000000
std      NaN      8.713647      36.649825      8.713647      3.464989

```

Figura 2. Resumen descriptivo, consumo total de agave.

b. Exportaciones por país

En este conjunto de datos se cuenta con una matriz de 64281 registros/filas y 9 columnas.

Columna	Tipo de Dato	Clasificación	Descripción
NombrePais	object	Categórica nominal	Nombre del país al que se exporta el producto.
Total_Pais_Mes	float64	Numérica continua	Valor monetario mensual total por país.
Categoria	object	Categórica nominal	Categoría del producto exportado.
Total_Categoria_Mes	float64	Numérica continua	Valor monetario mensual total por categoría.
Clase	object	Categórica ordinal	Tipo o maduración del producto (ej. Blanco, Joven, Reposado, Añejo, Extra añejo).
Litros 40 % Alc. Vol	float64	(sin clasificar)	Volumen de exportación por millones de litros
AñoArchivo	int64	(sin clasificar)	Periodo-Año correspondiente a la exportación.
Mes	int64	(sin clasificar)	Periodo-Mes correspondiente a la exportación.

Tabla 2. Diccionario de datos de exportaciones por país.

De este conjunto de datos se obtuvieron los siguientes resúmenes descriptivos:

Resumen descriptivo:

	NombrePaís	Total_Pais_Mes	Categoria \
count	64281	6.428100e+04	64281
unique	148	NaN	2
top	ESTADOS UNIDOS DE AMERICA	NaN	TEQUILA 100% DE AGAVE
freq	2782	NaN	32280
mean	NaN	6.265878e+05	NaN
min	NaN	4.000000e-01	NaN
25%	NaN	1.075905e+04	NaN
50%	NaN	3.448638e+04	NaN
75%	NaN	1.094237e+05	NaN
max	NaN	3.394919e+07	NaN
std	NaN	2.980709e+06	NaN

	Total_Categoria_Mes	Clase	Litros 40 % Alc. Vol	AnioArchivo \
count	6.428100e+04	64281	6.428100e+04	64281.000000
unique	NaN	5	NaN	NaN
top	NaN	BLANCO	NaN	NaN
freq	NaN	21471	NaN	NaN
mean	3.229436e+05	NaN	8.031691e+04	2013.213376
min	3.000000e-01	NaN	0.000000e+00	1997.000000
25%	4.389000e+03	NaN	7.401600e+02	2008.000000
50%	1.319976e+04	NaN	3.841500e+03	2014.000000
75%	5.029015e+04	NaN	1.397250e+04	2020.000000
max	2.554612e+07	NaN	1.402942e+07	2025.000000
std	1.660792e+06	NaN	5.515349e+05	7.502797

	Mes	Fecha
count	64281.000000	64281
unique	NaN	NaN
top	NaN	NaN
freq	NaN	NaN
mean	6.567524	2013-09-04 07:41:02.892612352
min	1.000000	1997-01-01 00:00:00
25%	4.000000	2008-03-01 00:00:00
50%	7.000000	2014-06-01 00:00:00
75%	10.000000	2020-01-01 00:00:00
max	12.000000	2025-03-01 00:00:00
std	3.434762	NaN

Figura 3. Resumen descriptivo de exportaciones por país.

Este conjunto de datos, como se puede ver, presenta una alta cardinalidad comparada a otros conjuntos de datos a trabajar para este proyecto.

Cardinalidad	
NombrePaís	148
Categoria	2
Clase	5

Tabla 3. Cardinalidad de exportaciones por país.

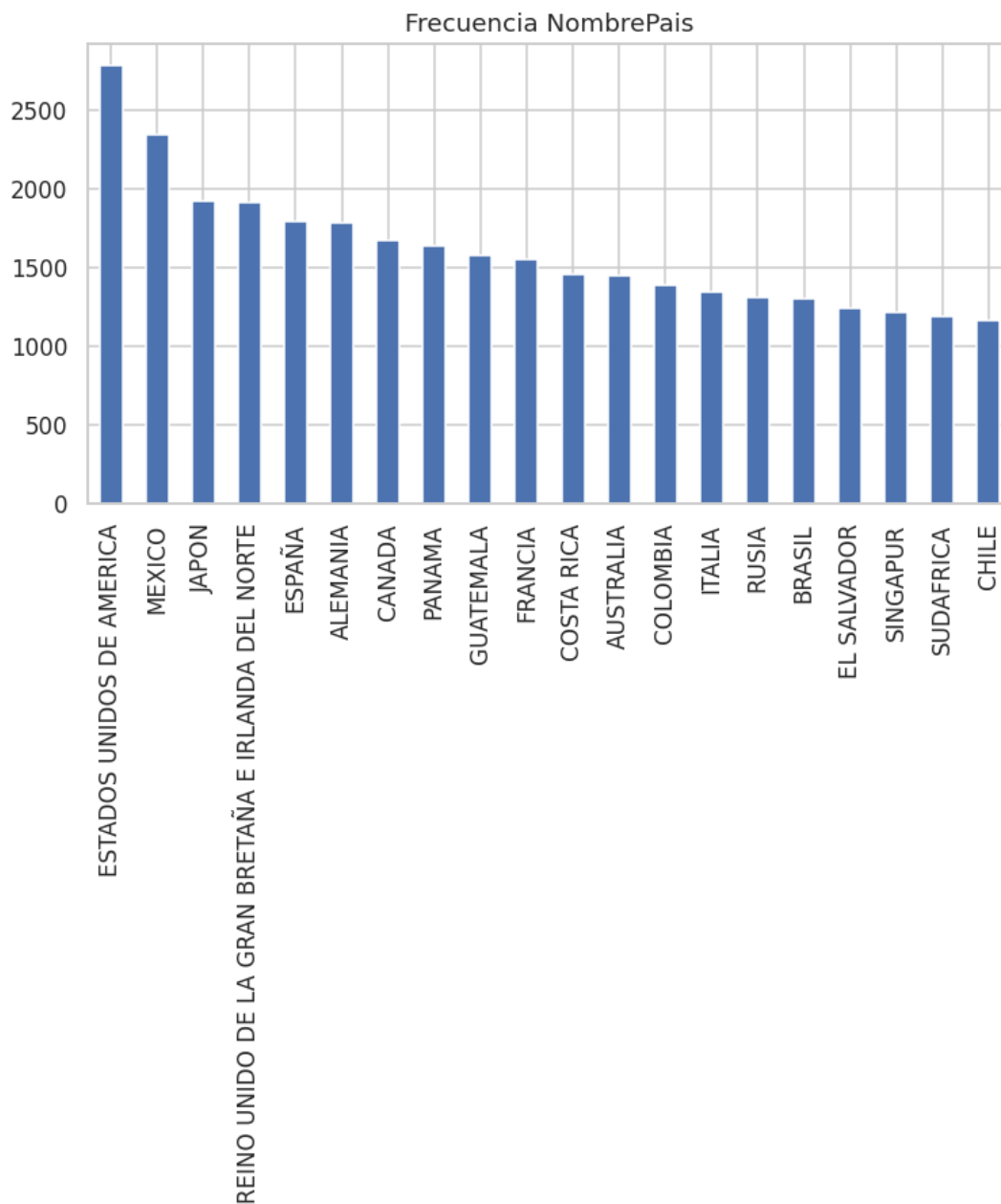


Figura 4. Frecuencia de valores en “NombrePaís” en exportaciones por país.

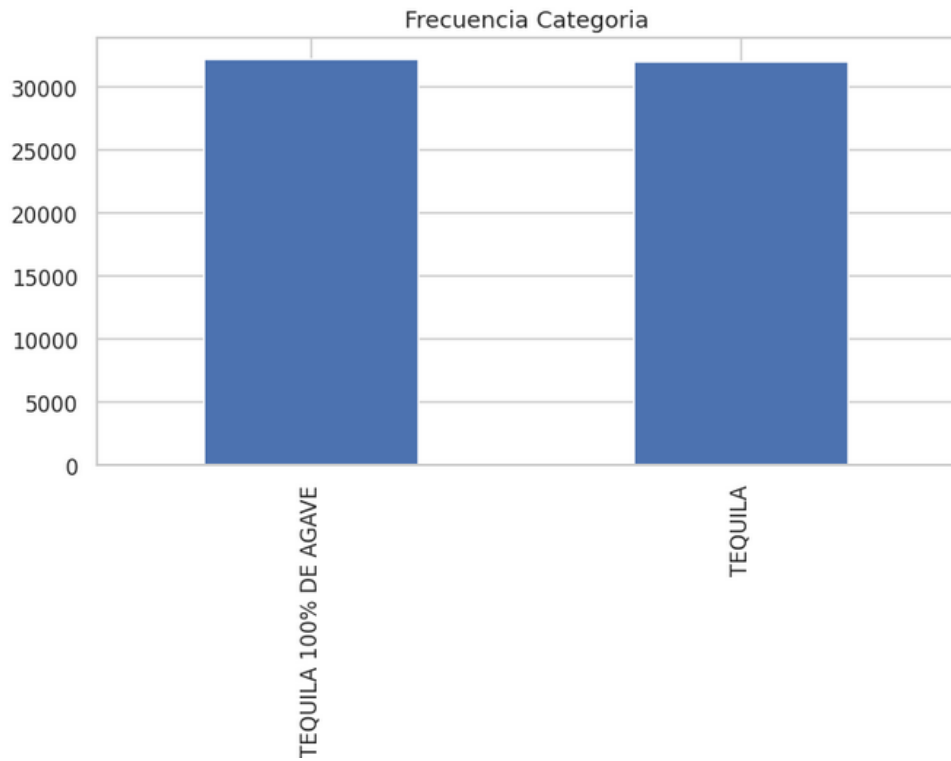


Figura 5. Frecuencia de valores en “Categoría” en exportaciones por país.

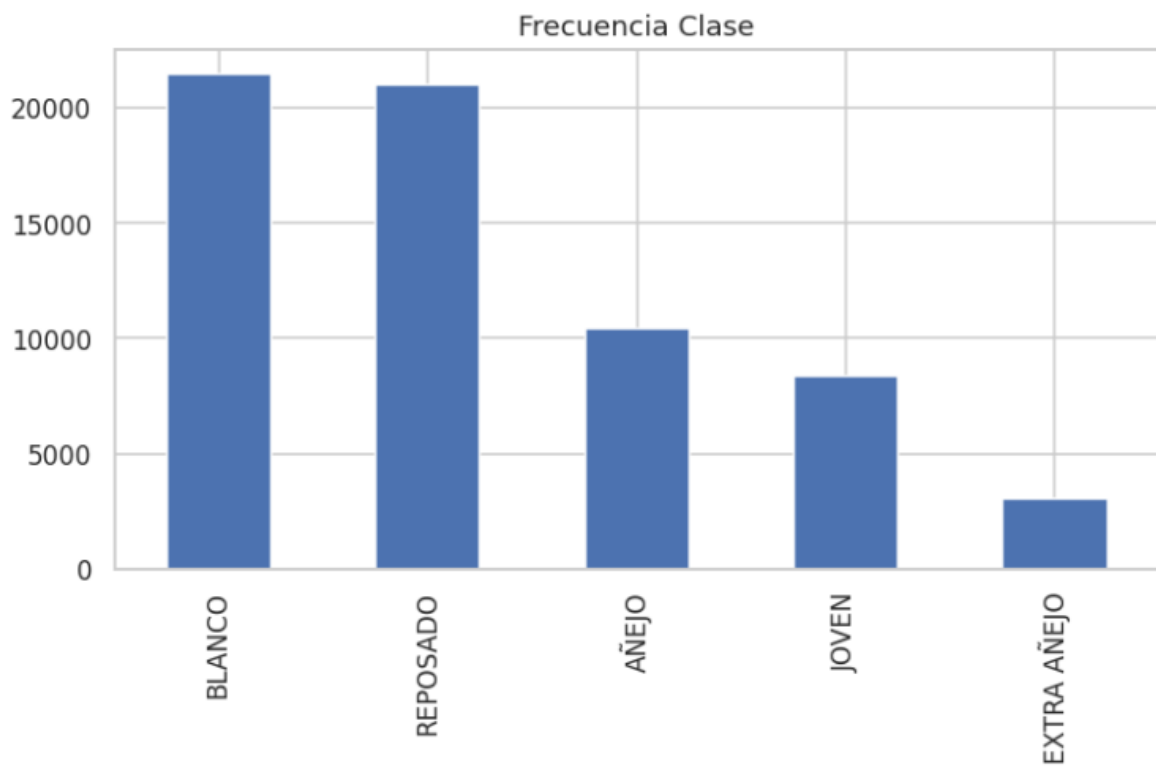


Figura 6. Frecuencia de valores en “Clase” en exportaciones por país.

c. Exportaciones Total de Forma

En este conjunto de datos se cuenta con una matriz de 726 registros/filas y 6 columnas.

Resumen descriptivo:

	SubCategoría	Year	Valor	AñoArchivo	Mes
count	726	726.000000	726.000000	726.000000	726.000000
unique	2	NaN	NaN	NaN	NaN
top	Envasado	NaN	NaN	NaN	NaN
freq	363	NaN	NaN	NaN	NaN
mean	NaN	2009.628099	7.431956	2009.628099	6.462810
min	NaN	1995.000000	0.400000	1995.000000	1.000000
25%	NaN	2002.000000	4.500000	2002.000000	3.000000
50%	NaN	2010.000000	6.500000	2010.000000	6.000000
75%	NaN	2017.000000	8.500000	2017.000000	9.000000
max	NaN	2025.000000	31.000000	2025.000000	12.000000
std	NaN	8.739099	5.620583	8.739099	3.464997

	Fecha
count	726
unique	NaN
top	NaN
freq	NaN
mean	2010-01-30 14:48:35.702479360
min	1995-01-01 00:00:00
25%	2002-07-08 18:00:00
50%	2010-02-01 00:00:00
75%	2017-08-24 06:00:00
max	2025-03-01 00:00:00
std	NaN

Figura 7. Resumen descriptivo de exportaciones totales por forma.

Columna	Tipo de Dato	Clasificación	Descripción
SubCategoría	object	Categórica nominal	Categoría del tequila usado para el consumo del agave (ej. Tequila 100%, Tequila).
Year	int64	Numérica continua	Año del registro de la exportación.
Valor	float64	Numérica	Exportación en millones de litros.
AñoArchivo	int64	Numérica continua	Año del archivado del registro de la exportación.
Mes	int64	Numérica continua	Mes donde fue exportado.

Tabla 4. Diccionario de datos de exportaciones por forma.

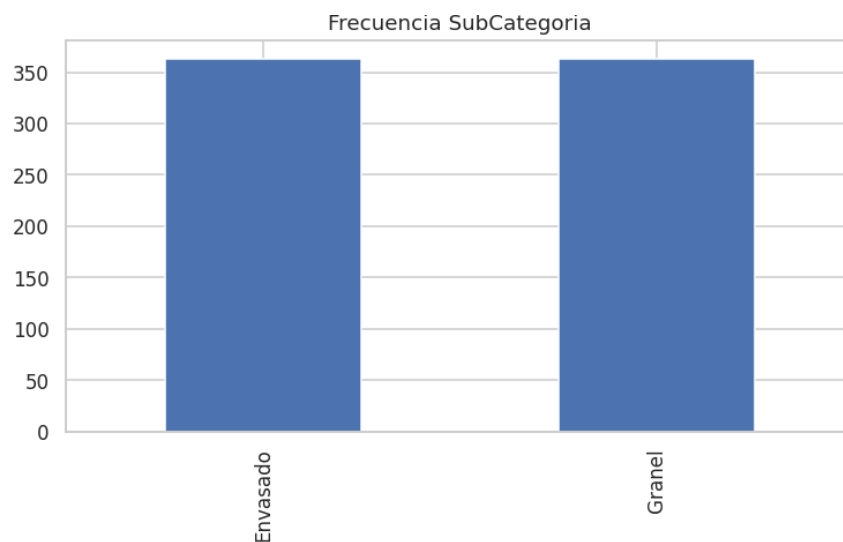


Figura 8. Frecuencia de valores de “Subcategoria” en exportación por forma.

d. Total de Exportaciones Categoría

En este conjunto de datos se cuenta con una matriz de 726 registros/filas y 6 columnas.

Columna	Tipo de Dato	Clasificación	Descripción
SubCategoria	object	Catagórica nominal	Categoría del tequila usado para el consumo del agave (ej. Tequila 100%, Tequila).
Year	int64	Numérica continua	Año del registro de la exportación.
Valor	float64	Numérica	Exportación en millones de litros.
AñoArchivo	int64	Numérica continua	Año del archivado del registro de la exportación.
Mes	int64	Numérica continua	Mes donde fue exportado.
Fecha	Date	Fecha	Fecha de exportación.

Tabla 5. Diccionario de datos de exportaciones totales por categoría.

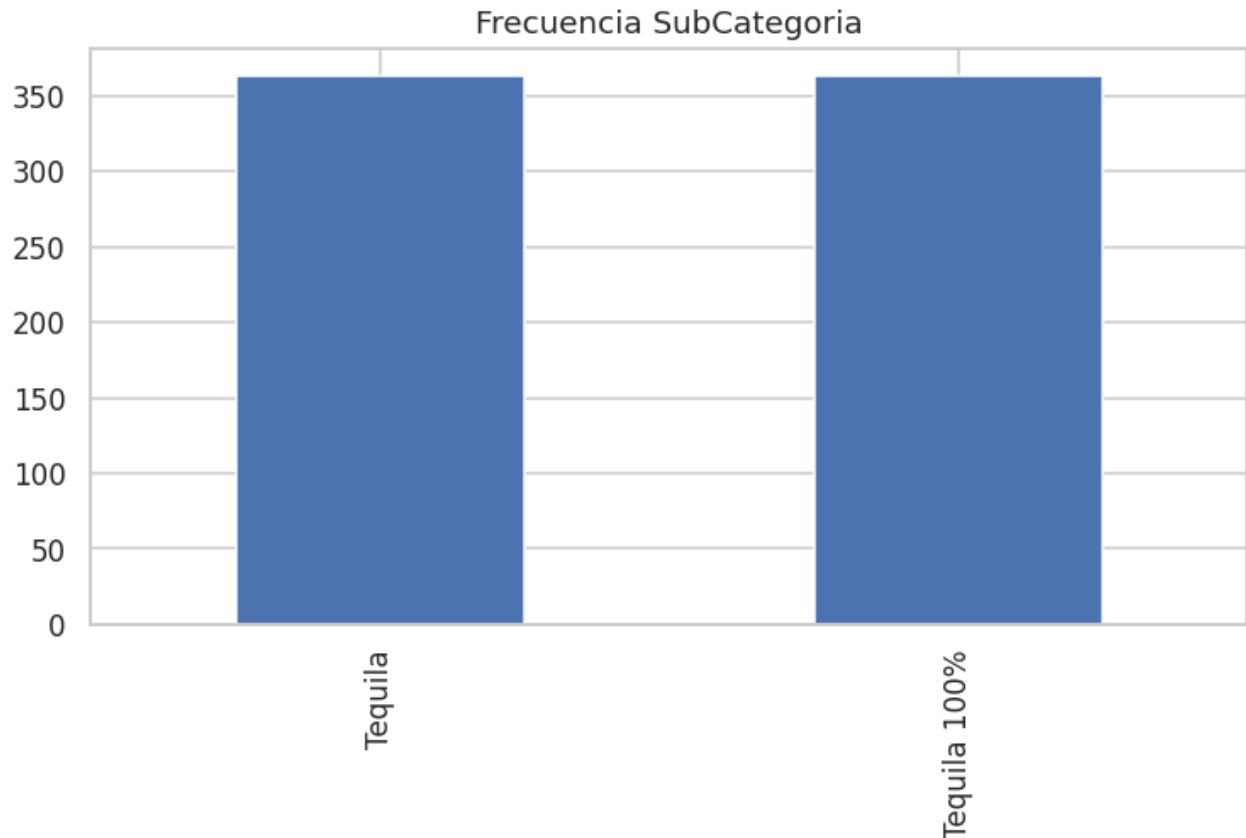


Figura 9. Frecuencia de clases en subcategoría, exportación total por categoría.

Preprocesamiento

Corrección de nulos

En los diferentes conjuntos de datos no se encontró ningún valor faltante, sin embargo, se agregó al pipeline de preprocesamiento un paso para la corrección de nulos en caso de que esto se llegase a presentar en algún futuro.

Winsorización de valores atípicos

Otro de los pasos del preprocesamiento realizado es la Winsorización, una transformación de datos a las columnas numéricas que estén fuera del 1.5 rango intercuartílico, pero se hace omisión de esta transformación en el caso de "Year", "Mes" y "AnioArchivo" debido a que estos datos a pesar de ser numéricos categóricos, no son cantidades.

Reducción de cardinalidad

En la gran mayoría de los conjuntos de datos se encontró que se tiene una baja cardinalidad en los valores categóricos. Sin embargo, se encontró que, en el conjunto de datos de exportaciones por país, la columna de “NombrePais” presenta una cardinalidad más alta comparado el resto de los valores categóricos.

Para reducir la cardinalidad se agrupan varios países con una frecuencia baja a un solo valor determinado como “OTROS”.

Corrección de sesgo

Debido a que las cantidades manejadas en estos conjuntos de datos presentan valores demasiado grandes, se procede a corregir el sesgo de las columnas con datos numéricos si presentan un sesgo mayor a 1 y si es así se le aplica la fórmula de logaritmo base 10 para disminuir el sesgo.

Borrado de datos innecesarios

En los distintos conjuntos de datos donde se usan las subcategorías o categorías del tequila s encuentran frecuentemente en los datos los valores de “Tequila 100% de Agave”, “Tequila”, “Total”.

El valor de Total es la suma de los valores de Tequila 100% y Tequila por lo cual no sirve de alguna manera analizar estos valores que son superfluos.

Descubrimientos

Valores faltantes / ausencia

Los conjuntos de datos “Total Exportaciones Forma”, “Total de Exportaciones Categoría”, “Producción Total Tequila”, “Consumo de Agave”, “Exportaciones País”, **no presentaron valores faltantes** en todos sus atributos.

a. Exportaciones Total Forma

Columna	% Nulos
SubCategoria	0%
Year	0%
Valor	0%
AnioArchivo	0%
Mes	0%
Fecha	0%

Tabla 6. Porcentaje de valores nulos.

b. Total de Exportaciones Categoría

Columna	% Nulos
SubCategoria	0%
Year	0%
Valor	0%
AnioArchivo	0%
Mes	0%
Fecha	0%

Tabla 7. Porcentaje de valores nulos.

c. Producción Total

Columna	% Nulos
SubCategoria	0%
Year	0%

Valor	0%
AnioArchivo	0%
Mes	0%
Fecha	0%

Tabla 8. Porcentaje de valores nulos.

d. Consumo de Agave Total

Columna	% Nulos
SubCategoria	0%
Year	0%
Valor	0%
AnioArchivo	0%
Mes	0%
Fecha	0%

Tabla 9. Porcentaje de valores nulos.

e. Exportaciones por País

Columna	% Nulos
NombrePais	0%
Total_Pais_Mes	0%
Categoria	0%
Total_Categoria_Mes	0%
Clase	0%
Litros 40 % Alc. Vol	0%
AnioArchivo	0%
Mes	0%
Fecha	0%

Tabla 10. Porcentaje de valores nulos.

Correlaciones

a. Exportaciones Total Forma

La máxima correlación en las variables numéricas se encuentra entre “Valor” y “Year” (0.72).

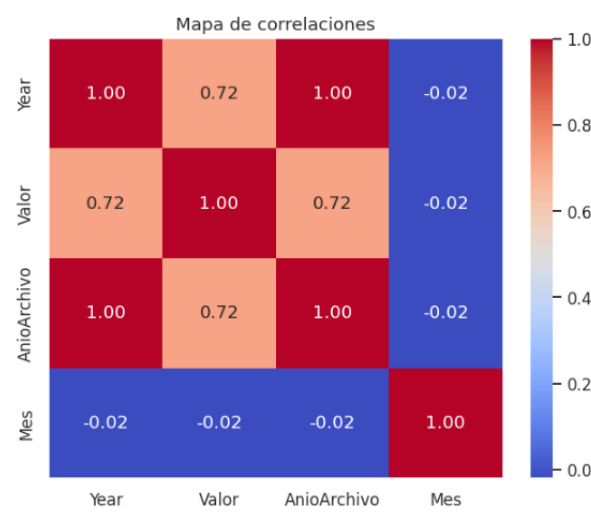


Figura 10. Mapa de calor, correlaciones de valores.

b. Total de Exportaciones Categoría

Tampoco se encontraron discrepancias para este dataset, pues las variables numéricas son las mismas, pero diferenciando la categoría: siendo la máxima correlación en las variables numéricas se encuentra entre “Valor” y “Year” (0.72).

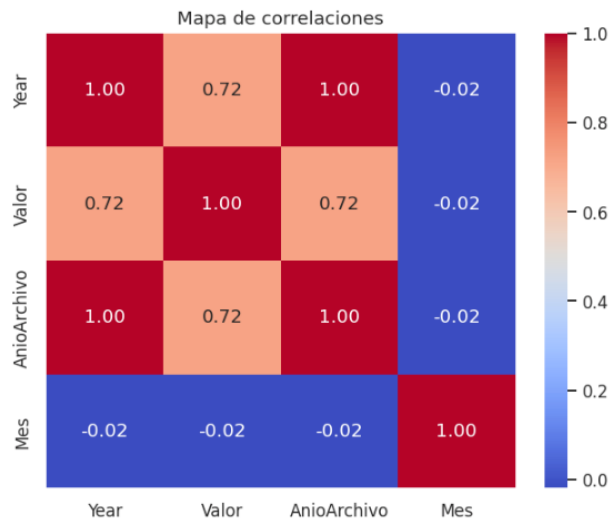


Figura 11. Mapa de calor, correlaciones de valores.

c. Producción Total

En la producción Total, las variables con mayor correlación son entre las variables numéricas se encuentra entre “Valor” y “Year” (0.62).

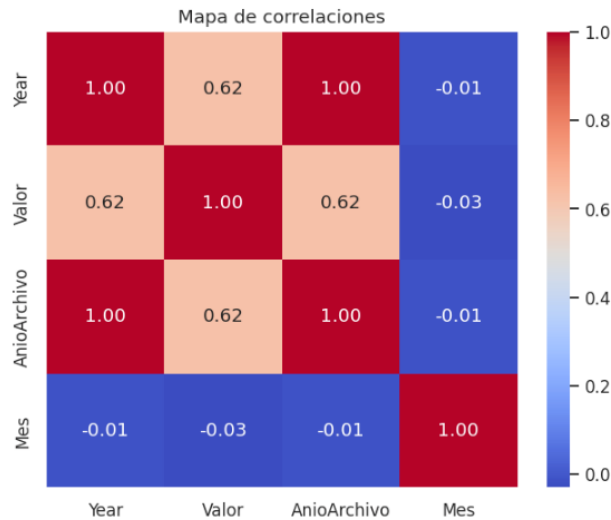


Figura 11. Mapa de calor, correlaciones de valores.

d. Consumo de Agave Total

Las variables con mayor correlación son entre las variables numéricas se encuentra entre “Valor” y “Year” (0.52).

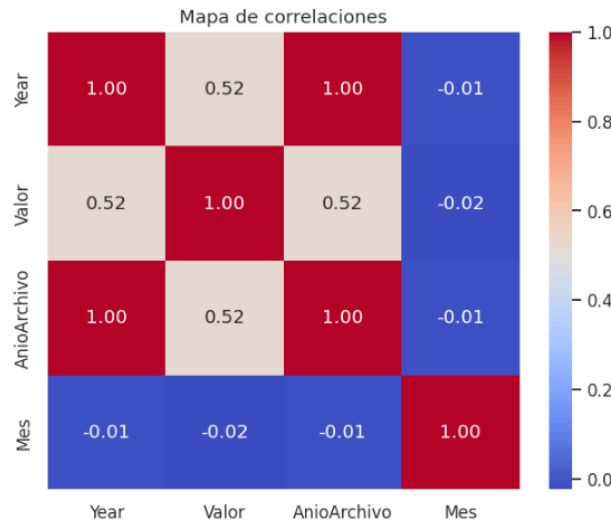


Figura 12. Mapa de calor, correlaciones de valores.

e. Exportaciones por País

Existen dos conjuntos de variables con correlación positiva:

- Total de Categoría mes y el Total país mes: 0.93.
- Total de Categoría mes y el Litros 40% vol. Alcohol : 0.66.

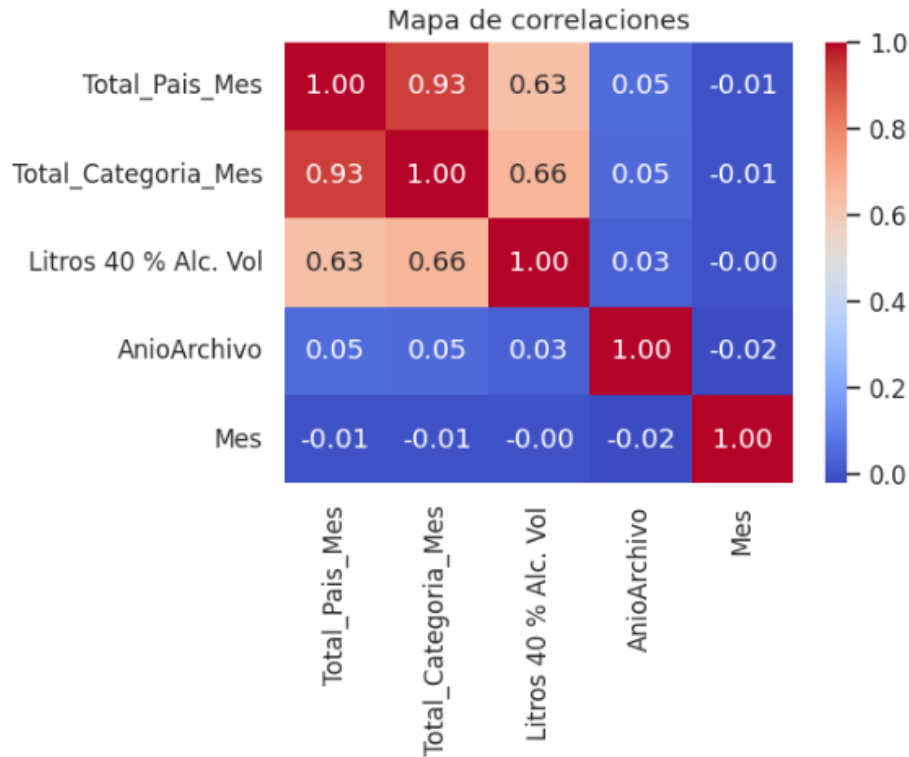


Figura 13. Mapa de calor, correlaciones de valores.

Valores atípicos

a. Exportaciones Total de Forma

El histograma correspondiente a la variable “Valor” tiene un **sesgo a la derecha** (asimétrica positiva); concentrando la mayor parte de los puntos entre los valores 3 y 10. Sin embargo, el boxplot ayuda a identificar valores extremos considerados como outliers superiores al valor de 14.5. La mediana es en torno al valor 6.5 lo que corresponde a la distribución.

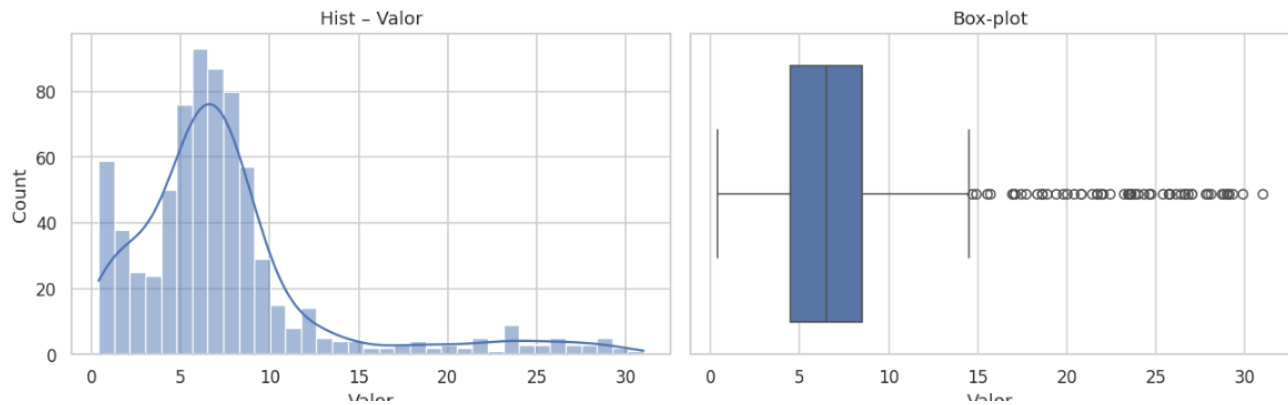


Figura 14. Distribución de la variable “valor” (izquierda) y boxplot (derecha) para el conjunto de datos “Exportaciones Total de Forma”.

b. Total de Exportaciones Categoría

La distribución también presenta un sesgo a la derecha para la variable “valor”, con una concentración de los valores entre los puntos 5 y 10. La mediana mostrada en el diagrama boxplot está cerca del valor 6.5; también existen valores atípicos, pero con una cardinalidad pequeña, respecto al conjunto de datos. La distribución presenta cierta simetría.

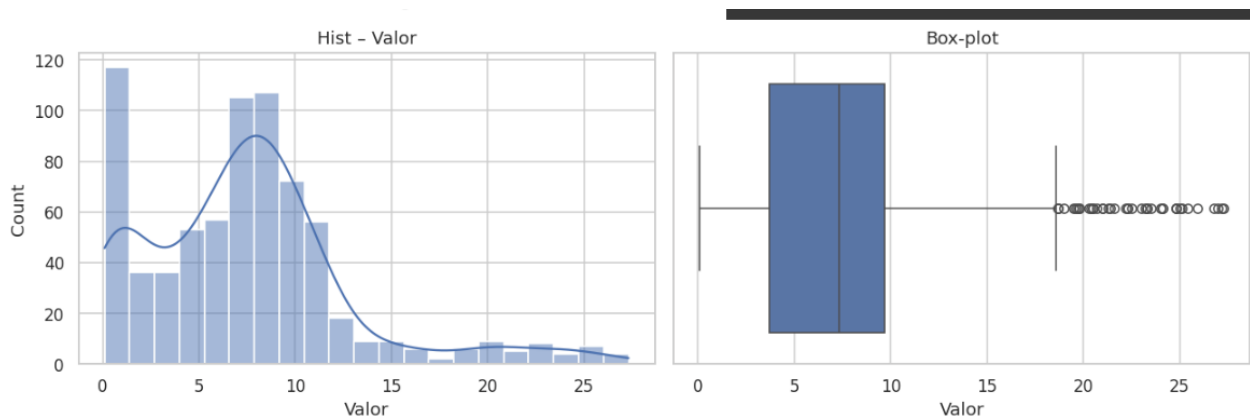


Figura 15. Distribución de la variable “valor” (izquierda) y boxplot (derecha) para el conjunto de datos “Total de Exportaciones por Categoría”.

c. Producción Total

La distribución tiene una mediana cerca al valor 10, con mayor concentración de los puntos entre el 7 y 10. El boxplot muestra la existencia de valores atípico que alcanzan el valor 48 y son un cúmulo importante de outliers. No es una distribución simétrica.

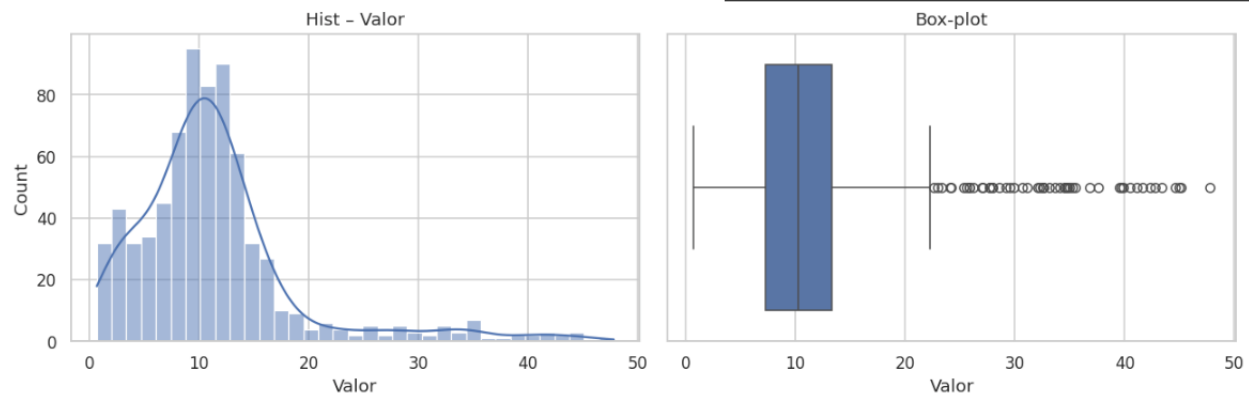


Figura 16. Distribución de la variable “valor” (izquierda) y boxplot (derecha) para el conjunto de datos “Producción Total”.

d. Consumo de Agave Total

La distribución muestra un sesgo a la derecha (asimetría positiva), los valores están centrados alrededor del punto 30, y la mayor parte de ellos se encuentra entre los puntos 20 y 40. Existen valores atípicos superiores al punto 80 pero con extremos hasta el valor 200.

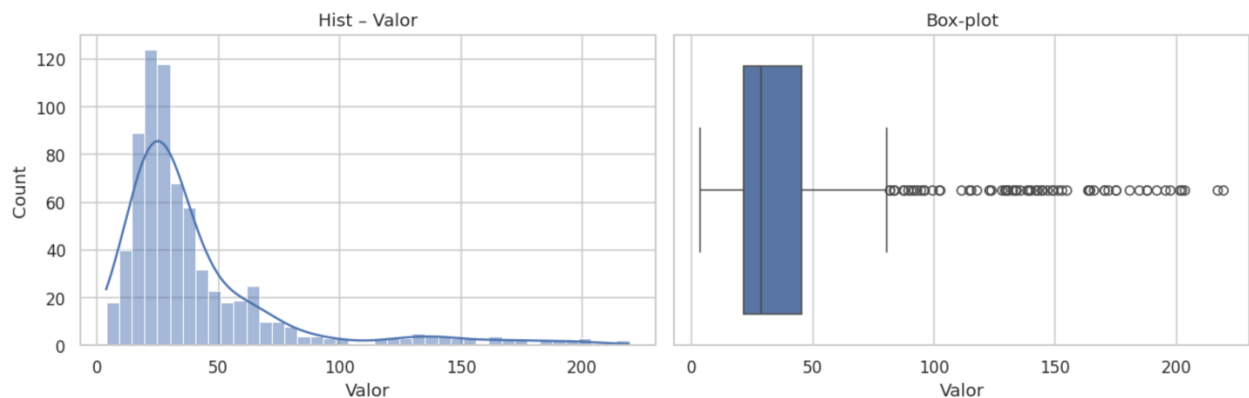


Figura 17. Distribución de la variable “valor” (izquierda) y boxplot (derecha) para el conjunto de datos “Consumo de Agave Total”.

e. Exportaciones por País

La distribución es sesgada a la derecha con una asimetría positiva extrema. Existen un gran volumen de ourliers dispersos. Esta es la variable con más asimetría de los 4 restantes.

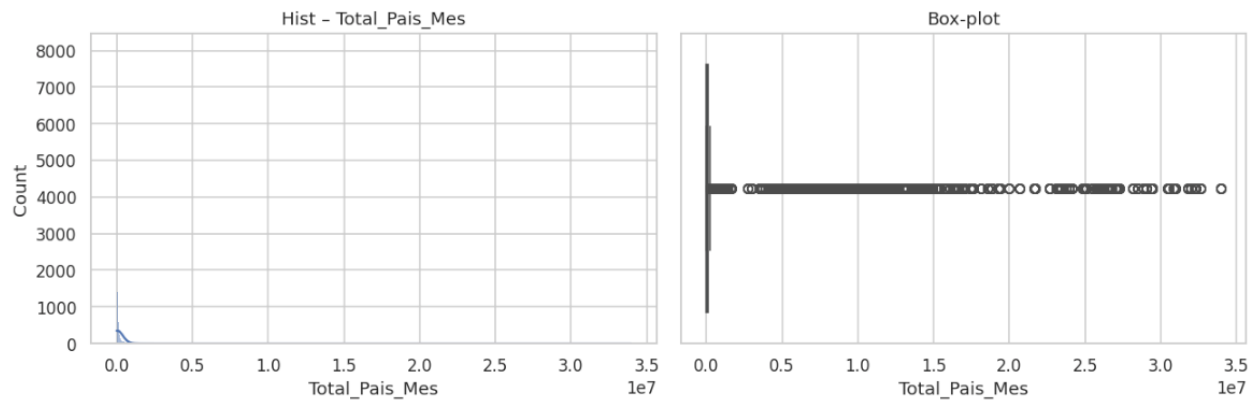


Figura 18. Distribución de la variable “valor” (izquierda) y boxplot (derecha) para el conjunto de datos “Exportaciones por País”.

Tendencias temporales

a. Exportaciones Total de Forma

Como se puede ver en la siguiente figura, existe una tendencia a exportar tequila envasado (línea naranja), mientras que la exportación a granel a pesar de su volatilidad tiene una tendencia de aumento más baja (línea azul). La tendencia de exportación del tequila envasado tuvo un crecimiento acelerado a partir del año 2020; el tequila a granel se mantuvo estable durante el mismo periodo.

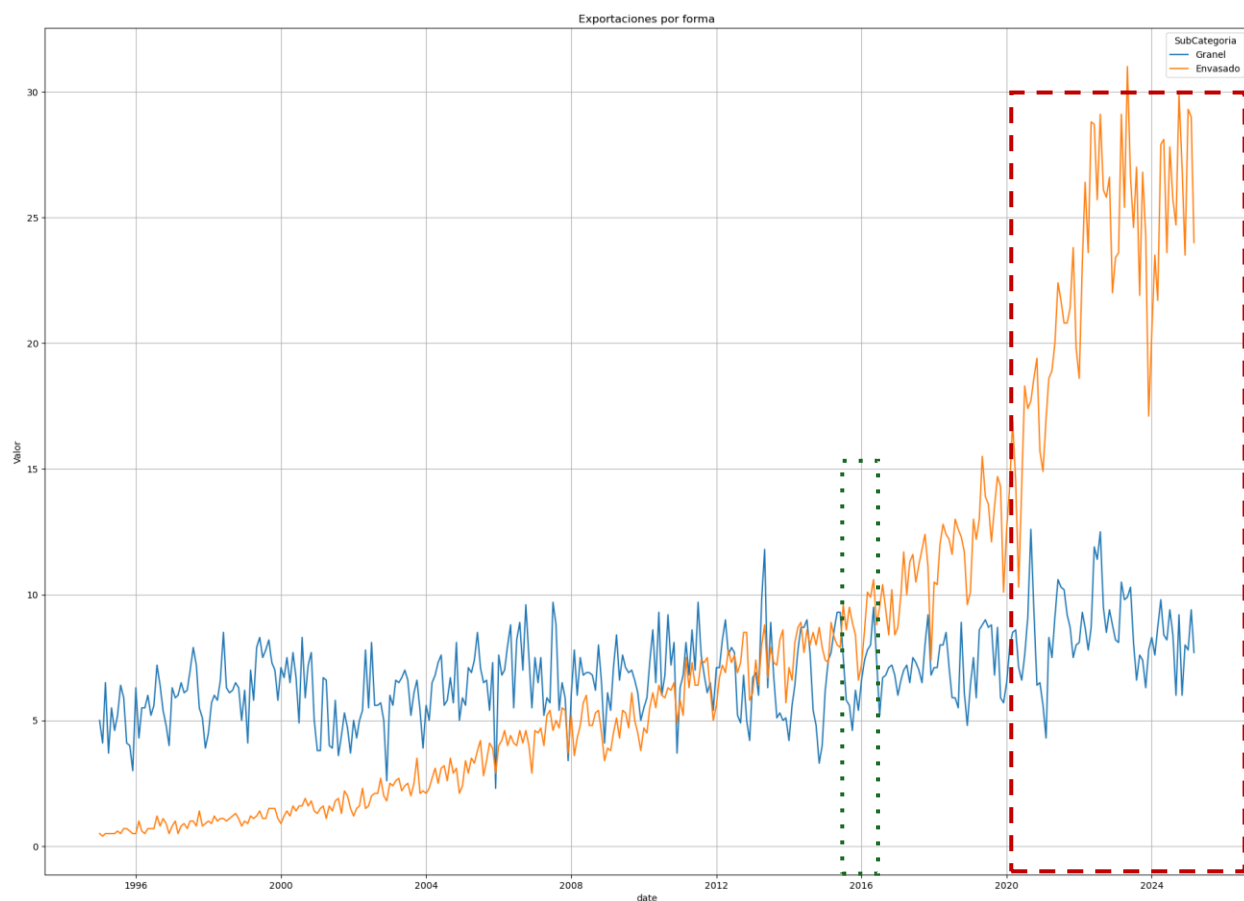


Figura 19. Exportaciones totales de tequila en forma, valor en millones de litros.

b. Total de Exportaciones Categoría

La tendencia del tequila 100% (línea azul) es creciente de manera general, sin embargo, en el año 2000 se observa un crecimiento con mayor aceleración respecto a los periodos anteriores, dicho crecimiento acelerado parece estabilizarse posterior al año 2024.

La tendencia del Tequila (línea naranja), es creciente, pero con una aceleración mucho más lenta; existen periodos oscilantes sin crecimiento entre los años 2008 y 2020. Ambas tendencias exhiben puntos estacionales cercanos a los inicios o términos de cada año.

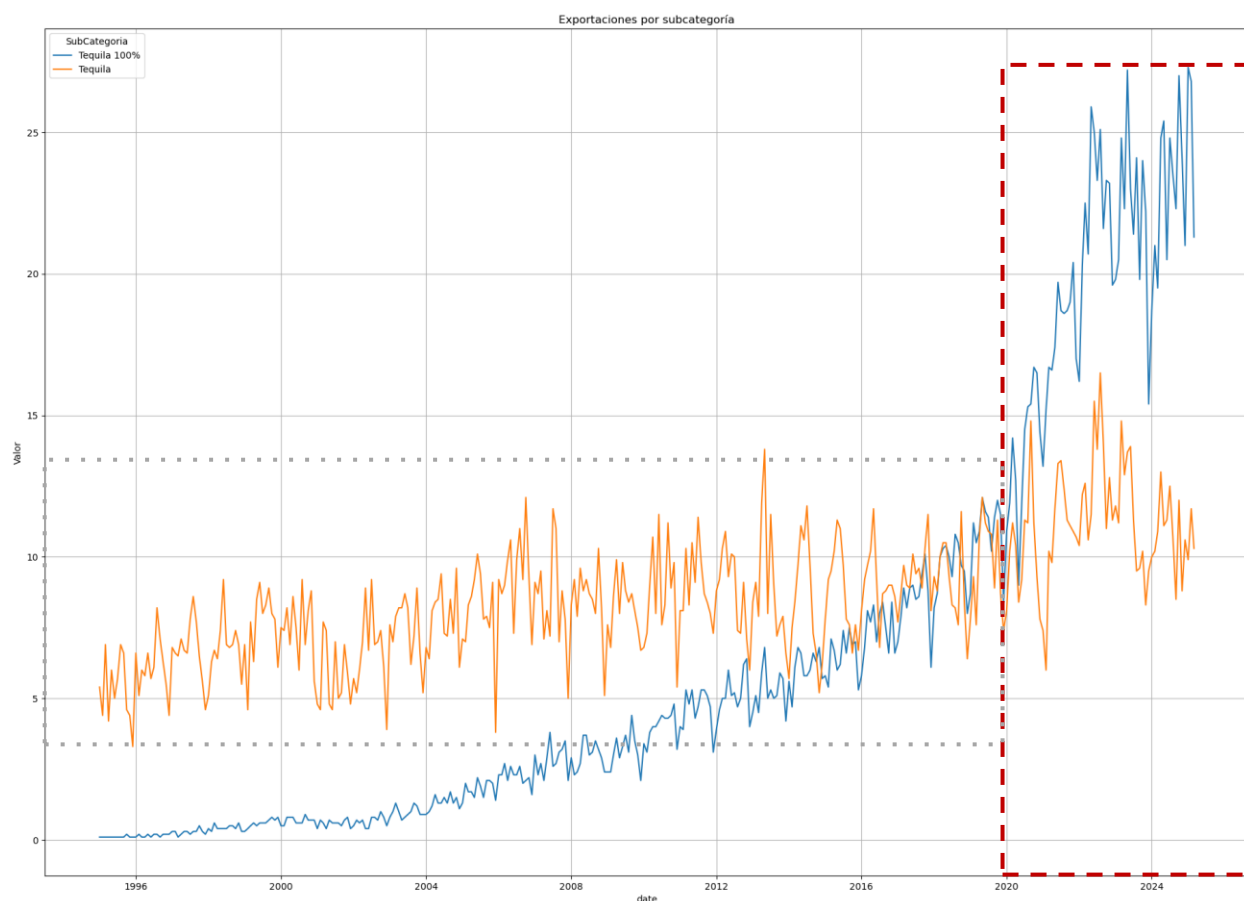


Figura 20. Exportaciones de tequila por subcategoría, valor en millones de litros.

c. Producción Total

La tendencia de la producción del Tequila 100% (línea azul) presenta tres periodos de crecimiento marcados: desde el comienzo hasta el año 2006, posterior al 2006 y posterior al 2022 y un periodo de aceleración creciente posterior al 2022. En el periodo intermedio se muestra crecimientos oscilantes en donde el Tequila (línea naranja) supera la producción con una distancia entre puntos no tan distantes.

En el periodo inicial y final la producción del Tequila 100% respecto al Tequila siempre se encuentra por arriba. Sólo en el tercer periodo existe un despunte del Tequila 100% de dos órdenes de magnitud.

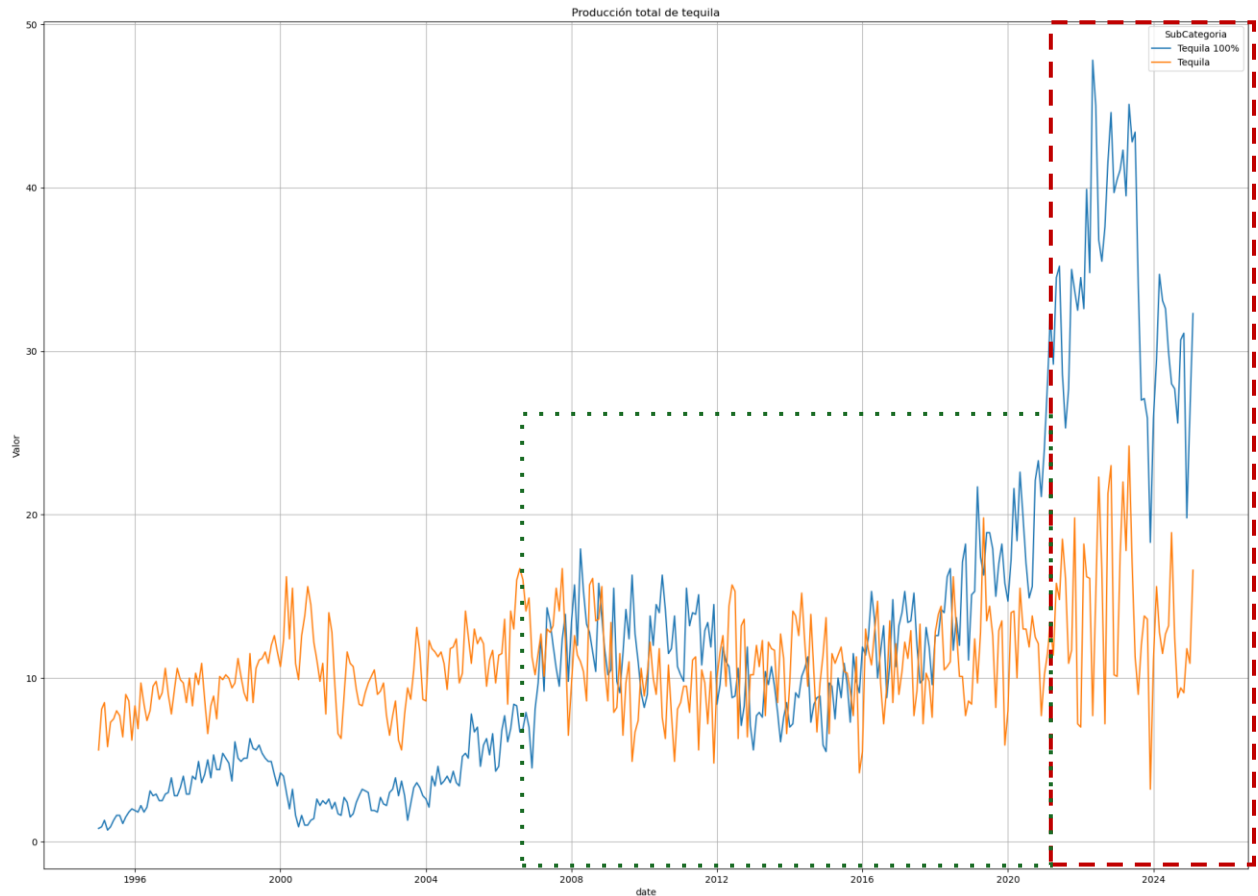


Figura 21. Producción total de tequila, valor en millones de litros.

d. Consumo de Agave Total

La tendencia del consumo de agave Tequila 100% (línea azul, figura 22) tiene periodos oscilantes con punto de inflexión (marcos punteados verdes) que tienen cierta temporalidad. Después del año 2022 existe un crecimiento acelerado en el consumo del agave, sin embargo, también tiene un punto de inflexión mostrando crestas y valles. La tendencia del consumo de agave Tequila (línea naranja), muestra una estabilidad oscilante, pero siempre acotada por una franja de consumo (marco morado)

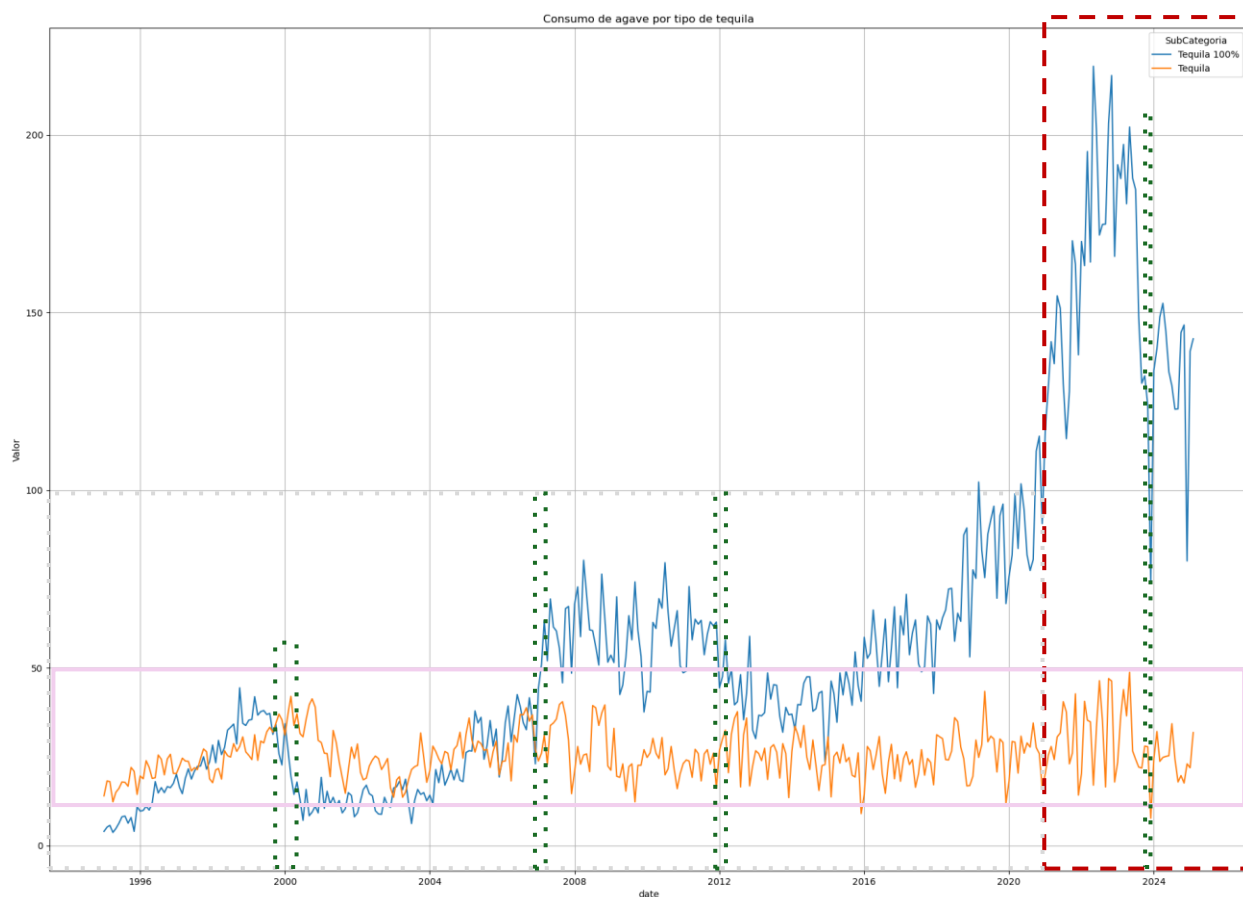


Figura 22. Consumo de agave por producción de tequila.

Las tendencias del consumo de agave (línea naranja) versus producción (línea azul) del Tequila 100%, se muestran en la figura 23; existe cierta estacionalidad en el desarrollo de ambas tendencias de manera sincronizada en periodos semejantes. Sin embargo, hay un periodo disruptivo a partir del año 2020 en donde el consumo muestra un crecimiento acelerado, con tres órdenes de magnitud respecto a la tendencia de producción, la estacionalidad aún está presente en ambas líneas de tendencia.

En la figura 24 sobre el consumo de agave versus la producción para el Tequila, se muestran ciertas estacionalidades en los mismos periodos; sin embargo, no hay puntos de inflexión en donde exista un crecimiento acelerado para ambas tendencias, lo que muestra cierta estabilidad.

Para ambos tipos de Tequila, el crecimiento sincrónico entre el consumo y la producción se visualiza en la figura 25, en donde hay cierta correlación positiva entre ambas variables.

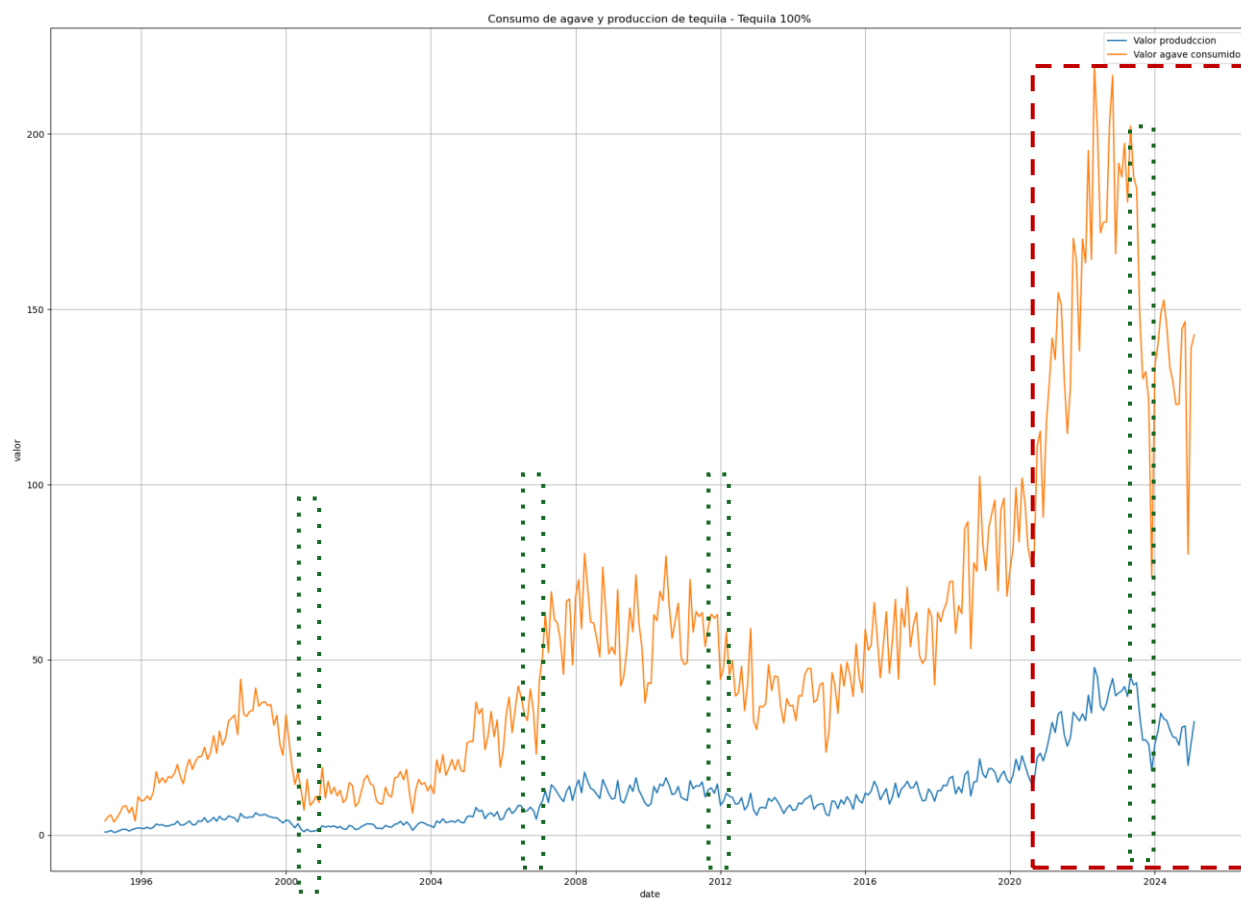


Figura 23. Consumo de agave y producción de tequila para Tequila 100%. Valor producción en millones de litros, valor en consumo en miles de toneladas.

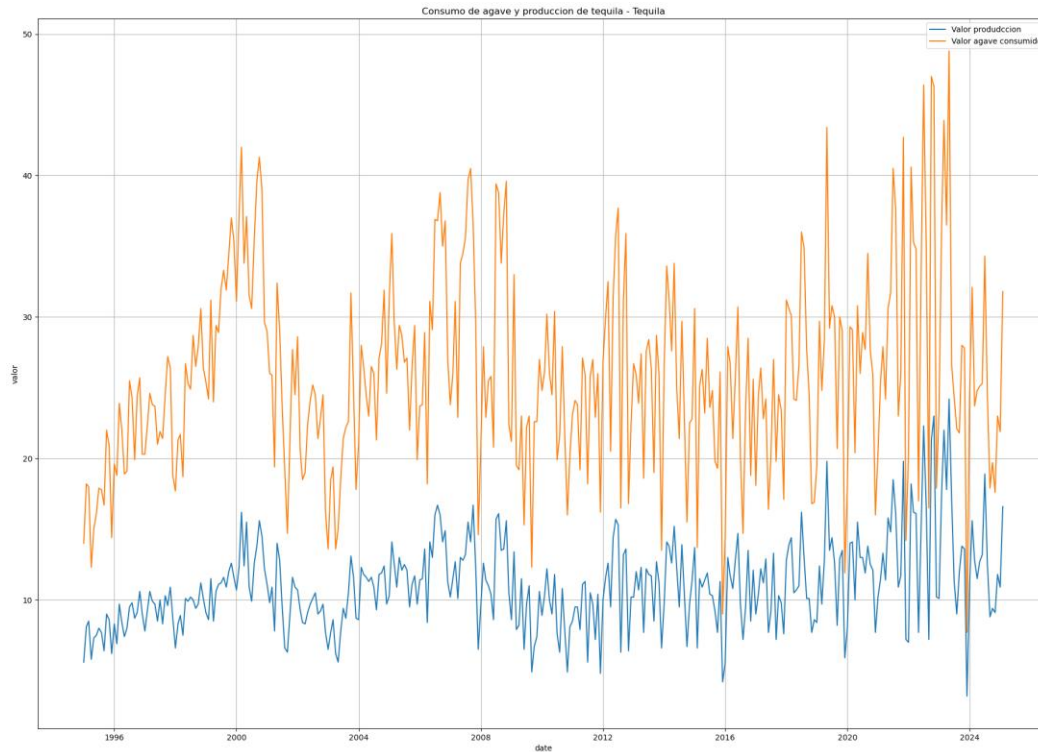


Figura 24. Consumo de agave y producción de tequila para Tequila. Valor producción en millones de litros, valor en consumo en miles de toneladas.

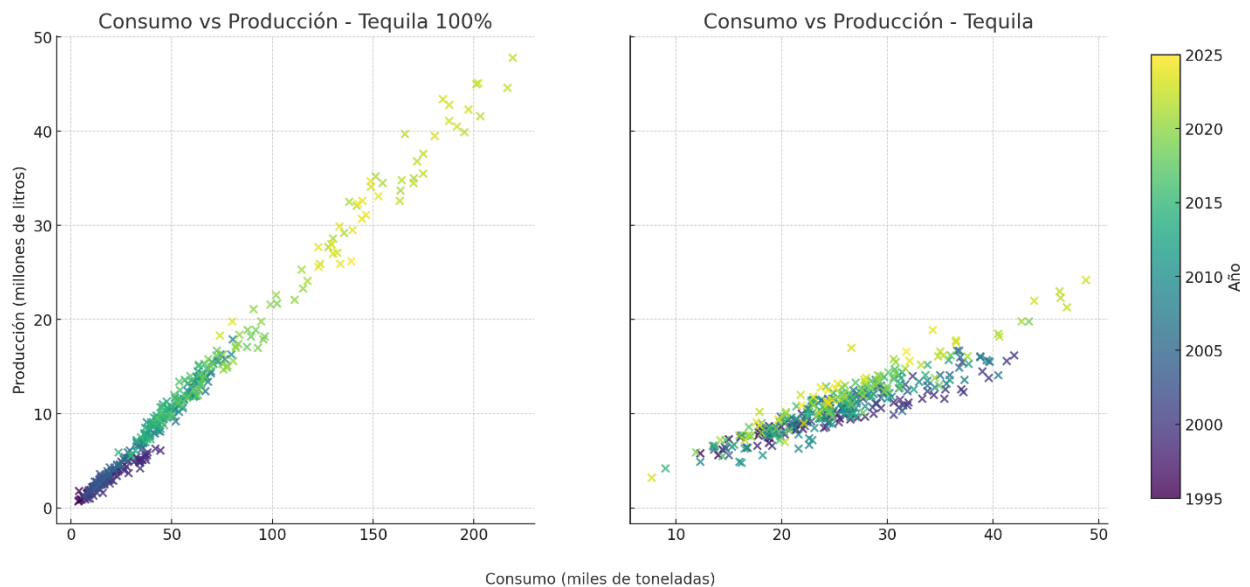


Figura 25. Gráfico de dispersión entre el consumo de agave y producción de tequila para Tequila 100% (izquierda) y Tequila (derecha). Valor producción en millones de litros, valor en consumo en miles de toneladas.

e. Exportaciones por País

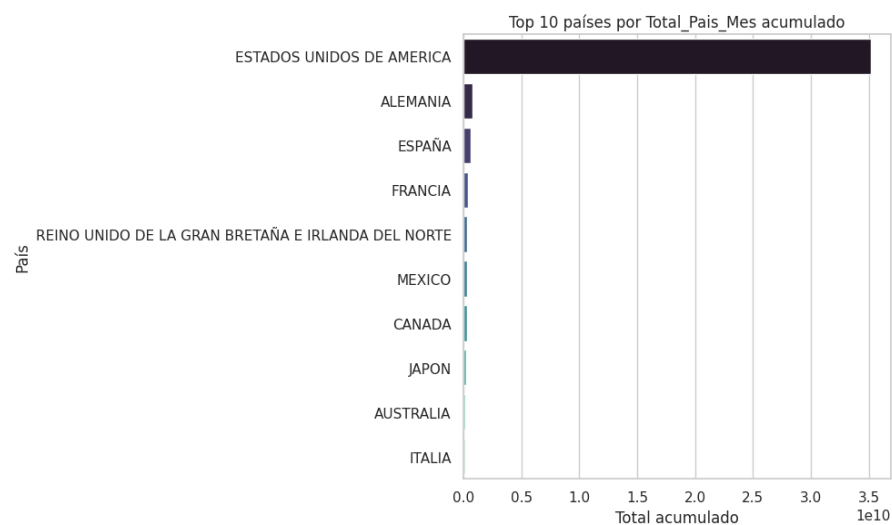


Figura 26. Top 10 países por total acumulado

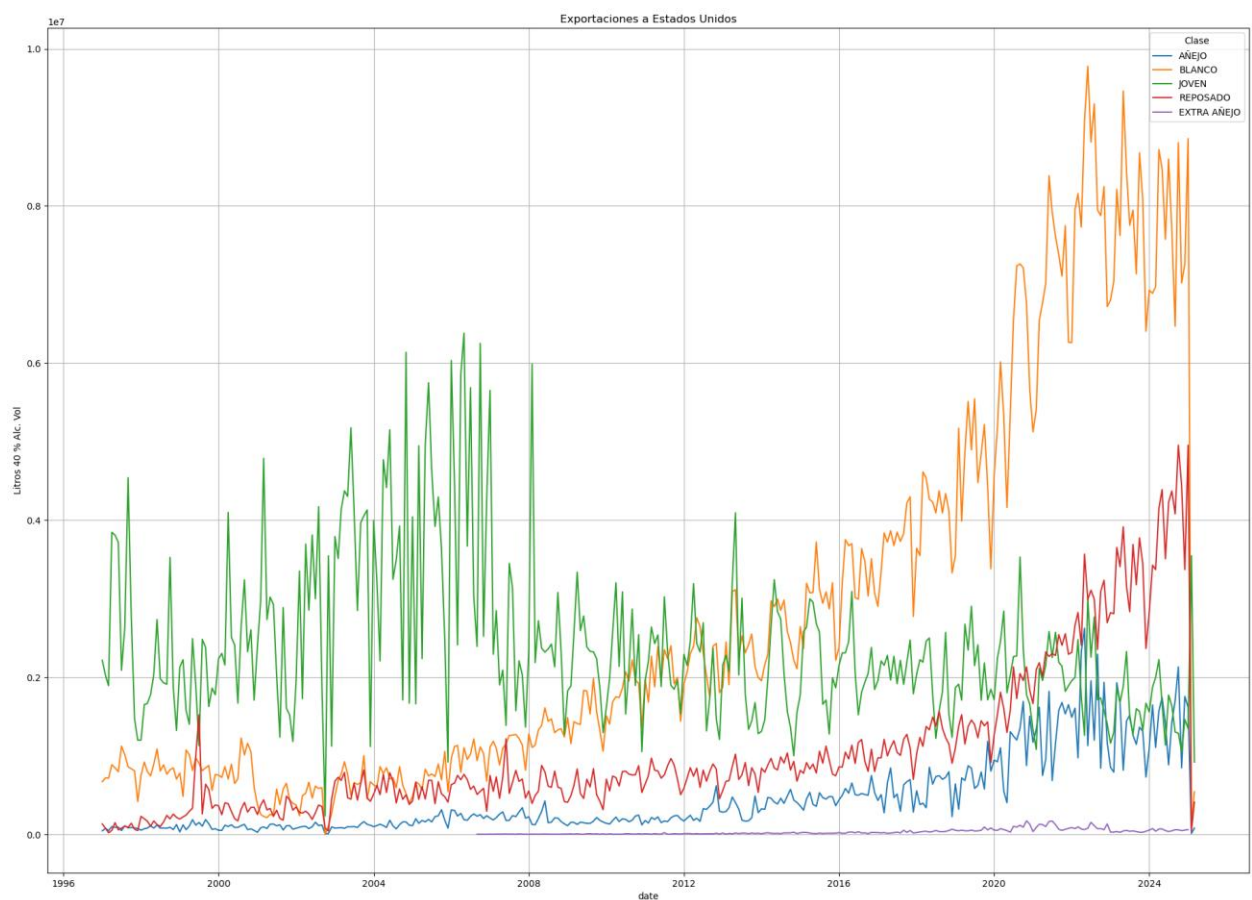


Figura 27. Total de tequila exportado por clase.

Desequilibrios

Como se ha visto en gráficas pasadas, la gran mayoría de nuestras variables categóricas se encuentran balanceadas ya que en los reportes obtenidos se encuentra un registro tanto de las categorías de Tequila 100% y Tequila.

Sin embargo, se encuentra que la variable categórica con más desequilibrio es “NombrePaís” donde se encuentra a Estados Unidos de América como la clase más predominante en este conjunto de datos.

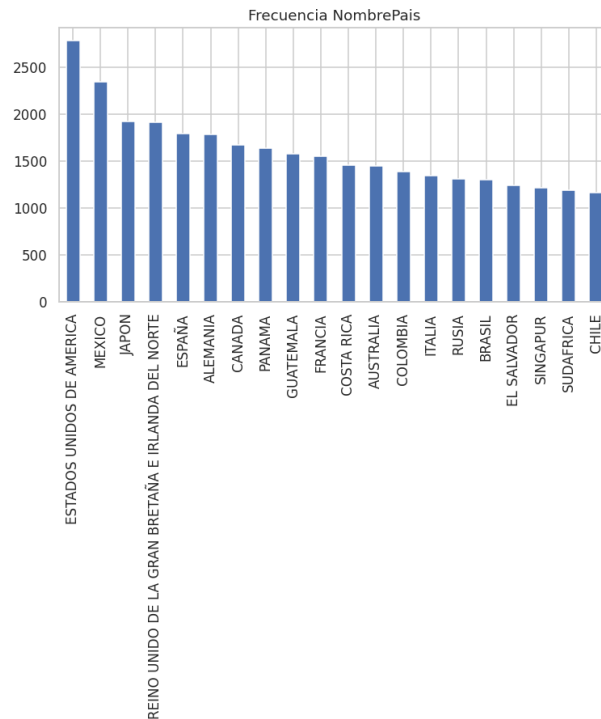


Figura 28. Frecuencia de valores en “NombrePaís” de las exportaciones totales por país.

Otra variable categórica donde se encuentra un desequilibrio es en los valores de “Clase” en las exportaciones totales por país, donde se encuentran los diferentes tipos de tequilas exportados, aquí se puede ver que los valores Añejo, Joven y Extra añejo, lo cual en el caso de añejo y extra añejo, hace sentido debido a la cantidad de tiempo necesario para su producción.

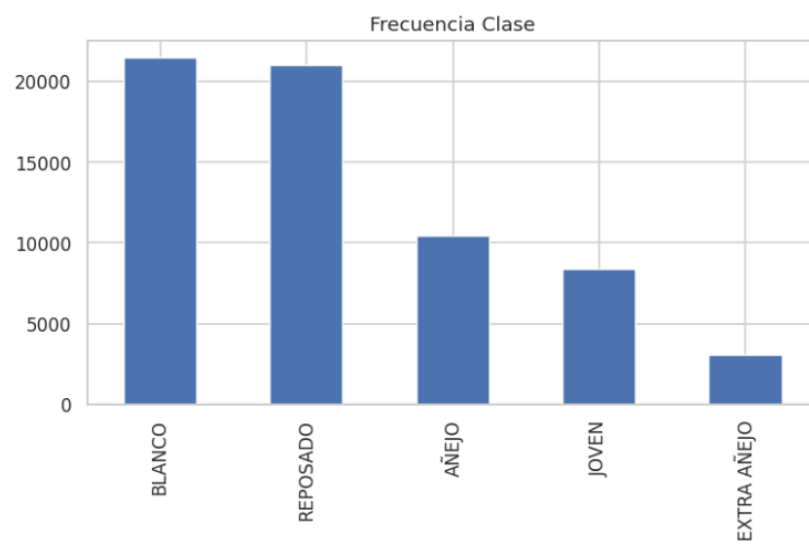


Figura 29. Frecuencia de valores en “Clase” de las exportaciones totales por país.

Conclusión

El análisis exploratorio de datos (EDA) permitió obtener una visión profunda del conjunto de datos y ha revelado varios aspectos cruciales para la continuación del proyecto:

1. Ausencia de valores faltantes: No se identificaron valores nulos en el conjunto de datos, lo que facilita el proceso de modelado, ya que no será necesario realizar imputaciones de datos faltantes.
2. Desafíos del modelo de series temporales: El análisis ha revelado que el conjunto de datos está estructurado temporalmente, lo que implica que el proyecto debe abordar un problema de predicción de series de tiempo (time series forecasting). Esto obliga a considerar adecuadamente la temporalidad, la estacionalidad y las posibles tendencias a lo largo del tiempo.
3. Variable objetivo: En todos los conjuntos de datos, la variable objetivo es el "Valor", lo que hace que los modelos de regresión sean adecuados para este tipo de predicción. Esta uniformidad facilitará la comparación de resultados entre los diferentes modelos.
4. Tendencia disruptiva post-2020: Un hallazgo clave es la aceleración del crecimiento en las exportaciones y consumo de agave después del año 2020. Este periodo disruptivo será fundamental a la hora de ajustar las ventanas de tiempo para el modelo, ya que podría influir significativamente en las predicciones y debe ser considerado en la calibración del modelo.
5. Importancia de Estados Unidos en las exportaciones: El análisis mostró que Estados Unidos es el principal país receptor de las exportaciones de tequila, lo que destaca la importancia de modelar adecuadamente las exportaciones hacia este país en particular para mejorar la precisión de las predicciones.
6. Correlación significativa entre consumo de agave y producción de tequila: Se observó una fuerte correlación entre el consumo de agave y la producción de tequila, lo que sugiere que el modelo debe considerar tanto el consumo como la producción como variables clave para las predicciones de exportación.
7. Desbalance en las clases: A pesar de que el conjunto de datos muestra una distribución relativamente equilibrada, el modelo podría necesitar ajustes para manejar posibles desbalances en las clases, especialmente en términos de tipos de productos y categorías de exportación.

Con base en los hallazgos obtenidos, se puede concluir que el problema principal a resolver es la predicción de exportaciones y producción de tequila como una serie temporal. Este problema debe abordarse teniendo en cuenta las siguientes acciones:

- **Enriquecimiento de los datos:** Es crucial identificar variables exógenas (como factores económicos, climáticos o regulatorios) que puedan influir en las exportaciones y el consumo de agave. La inclusión de estas variables permitirá mejorar la precisión del modelo.
- **Exploración de ventanas de tiempo:** Será necesario definir una ventana temporal adecuada para el modelo, especialmente considerando el cambio acelerado posterior a 2020. Esto puede incluir la prueba de modelos con diferentes periodos de entrenamiento y evaluación para ajustar los parámetros del modelo.
- **Consideración de tendencias y estacionalidad:** Se debe incorporar la estacionalidad y las tendencias a largo plazo en el modelo de series temporales para capturar adecuadamente los patrones cíclicos y las fluctuaciones a lo largo del tiempo.

En resumen, el EDA ha proporcionado información valiosa para la ingeniería de características y la construcción del modelo predictivo, destacando áreas claves como la influencia de Estados Unidos, la correlación entre consumo y producción, y la necesidad de un enfoque adecuado de series temporales con ventanas de tiempo ajustadas.