

Proyecto final

Programación y análisis estadístico con Python, con Raquel Peña

Technology & Data

En este proyecto estudiarás un conjunto de datos con muestras de vino tinto del norte de Portugal. El objetivo es modelar la calidad del vino en base a pruebas fisicoquímicas de la Calidad del Vino Rojo Portugués, el cual te proporcionaremos. Para ello será importante el analizar los datos, descubrir los patrones, probar hipótesis y realizar una regresión lineal para predecir la calidad del vino, así como su verificación.

El objetivo es Modelar la calidad del vino tinto mediante una regresión lineal, planteamiento de hipótesis , así como el análisis exploratorio de los datos a través de diferentes técnicas estadísticas.

Conocimientos a utilizar:

Estadística descriptiva, análisis exploratorio mediante visualizaciones y estadísticos, pruebas de hipótesis y regresión lineal en un Google Colab con código python.

herramientas a utilizar:

- Google Colab con código python





● Avance 1: Exploración y visualización de datos

Para este avance de proyecto debes descargar la base de datos WineQuality_Red.csv que está contenida en la parte de los adjuntos del curso. Esta base de datos contiene información de evaluaciones de la calidad del vino y otras variables fisicoquímicas. Las variables dentro de la base de datos son:

- 1 - fixed acidity
- 2 - volatile acidity
- 3 - citric acid
- 4 - residual sugar
- 5 - chlorides
- 6 - free sulfur dioxide
- 7 - total sulfur dioxide
- 8 - density
- 9 - pH
- 10 - sulphates
- 11 - alcohol
- 12 - quality (calificación entre 0 and 10) Variable objetivo (basada en datos sensoriales):

En este primer avance, el objetivo es realizar el análisis exploratorio de los datos y visualización de datos. Para ello algunos de los siguientes puntos que serán útiles verificar son las siguientes:

Paso 1: Importar la base de datos en formato csv, guardarlo como un dataframe y visualizar los primeros 4 renglones de la base para darte una idea de los componentes de esta. Además de obtener la información del tipo de variables que contiene el dataframe pues te será de mucha ayuda para conocer el tipo de gráfico que le conviene más a cada una.



- ¿Qué tipos de variables contiene el dataframe?

Paso 2.0: Realizar un análisis descriptivo de la base de datos. Encontrar para todas las variables: promedio, desviación estándar, mediana y percentil 75%.

Paso 2.1: Realizar un análisis descriptivo de la base de datos. Calcular la moda para todas las variables.

- ¿Se puede calcular la moda para variables numéricas?

● **Avance 2: Pruebas de Hipótesis y Probabilidades.**

En este segundo avance de tu proyecto consiste en verificar mediante un test de normalidad la distribución de una variable estableciendo de manera correcta la hipótesis nula y la hipótesis alternativa. Además, calcularás la probabilidad de obtener una calificación específica en la base de datos usando una tabla normal estándar (z).

Paso 1:

En el avance anterior realizaste un histograma para cada una de las variables. ¿A partir de esta qué forma notaste que tiene la variable ph? Ahora se confirma si los datos extraídos para la variable de pH para la muestra de Calidad del Vino Tinto provienen de una distribución normal, esto mediante el test de Normalidad Shapiro-Wilk. Esto bajo un nivel de significancia alfa del 0.05 y supón que las observaciones de cada muestra son independientes e idénticamente distribuidas.

1. Establece la hipótesis nula y la alternativa.
2. ¿Cuál es el valor del test estadístico? (valor entre 0 y 1).



3. ¿Cuál es el valor del p-value para esta prueba bajo tu hipótesis? Imprime tu resultado con 6 dígitos de exactitud para que veas lo que sucede (print('p=%.{6}f' % (p))))
4. Explica si rechazas o aceptas la hipótesis nula H_0 al nivel de significancia $\alpha = 0.05$. ¿Qué concluyes sobre la hipótesis?
5. De manera práctica, ¿cómo podrías verificar si la distribución del pH sigue una distribución normal? ¿Podrías decir que sí proviene de una distribución normal?

Paso 2:

Estandariza la variable alcohol y dibuja su histograma con 20 bins.

Paso 3:

Obtener la probabilidad de obtener un nivel de alcohol menor a 10 en la base de datos de la Calidad del Vino Tinto. Esta probabilidad la puedes obtener mediante el uso de la función en Python. ¡No olvides de usar el valor z (estandarizado)!

- ¿Qué puedes sugerir de los niveles de alcohol tomando como referencia la probabilidad anterior?

● Avance 3: Regresión lineal simple

Finalmente has llegado a la parte 3 y última de tu proyecto final en el que construirás un modelo de regresión simple para predecir la calidad (quality) del vino tomando como variable explicativa aquella que en tu análisis exploratorio haya tenido una correlación mayor con la variable dependiente.

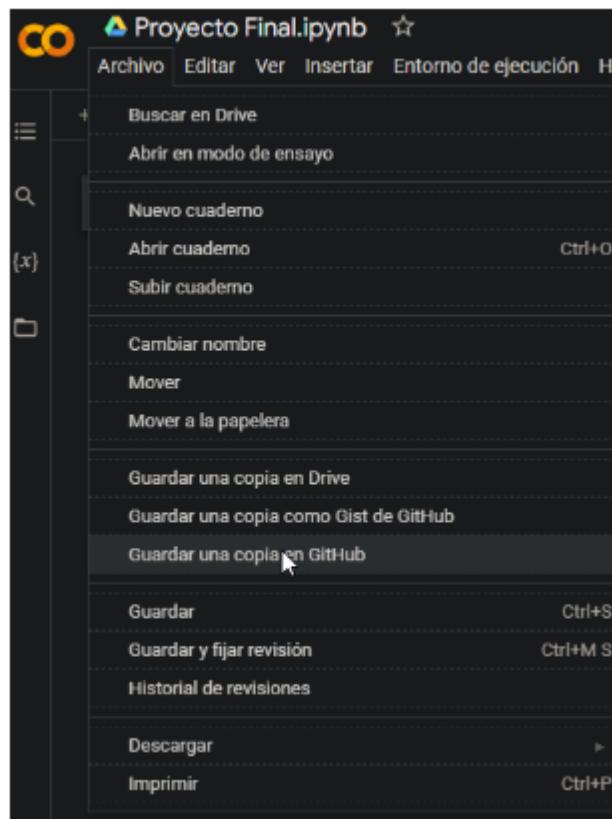


- Parte 1: Elige como variable explicativa aquella que estuvo más correlacionada con la calidad en la base de datos y como variable respuesta u objetivo la calidad.
- Parte 2: Separa en entrenamiento y validación tu prueba 80-20. Construye tu predicción de la regresión lineal simple sobre el conjunto de entrenamiento.
- Parte 3: Realiza dos visualizaciones de la dispersión de datos y la línea de ajuste tanto en el conjunto de entrenamiento como en el de validación.
- Parte 4: ¿Cuáles son los valores de las betas de la ecuación que describe la regresión lineal?
- Parte 5: Calcula las métricas de RMSE y la r-cuadrada tanto para el training y test.



- ¿Qué puedes concluir sobre este modelo de regresión?

Una vez que tengas completo el proyecto final guárdalo en GitHub, puedes descargar tu archivo desde el menú Archivo o bien guardar desde ese mismo menú directamente a GitHub



Recuerda que para que podamos retroalimentarte es necesario que nos compartas la liga de tu proyecto Final.

¡Éxito!



- Expectativas:

¿Tienes claridad con respecto a tu cargo y responsabilidades?

¿Tienes claridad con respecto a tus objetivos de corto, mediano y largo plazo?

- Fortalezas:

¿En tu trabajo tienes la oportunidad de utilizar tus fortalezas todos los días?

¿Conoces tus fortalezas y las fortalezas de tu equipo?

- Desarrollo:

¿Has recibido feedback significativo en los últimos 3 meses?

¿Has tenido oportunidad de crecimiento / desarrollo en los últimos 3 meses?

- Opiniones:

¿Tus opiniones son consideradas en el día a día laboral?

¿Estás involucrado en las tomas de decisión de tu área?

¿Sientes que puedes expresar tu opinión sin miedo?

- Misión y propósito:

¿La misión de (nombre de la empresa) te inspira y te hace sentir que tu trabajo es importante?

¿Tienes claridad de tu propósito / del propósito de tu rol dentro de la organización?

2.2. Incluye las preguntas adicionales que consideres pertinente.

2.3. Termina de estructurar tu encuesta:

- Incluye una sesión de bienvenida, explicando los objetivos de la encuesta
- Define si la vas a dividir por parte o sesiones
- Define el orden de las preguntas
- Define cuáles serán los tipos de respuestas: opción múltiple, escalas, ranking, respuesta cerrada o abierta, etc.

Tip: Cuánto más corta y sencilla tú encuesta, mejor.

Coloca aquí el enlace a tu encuesta:

<https://forms.gle/afZGK79xykBzRz1BA>



2.3. Utiliza la encuesta con tu equipo de trabajo (Opcional)

La idea es que puedas utilizar esta encuesta para evaluar los niveles de engagement de tu equipo de trabajo, por si tu empresa no cuenta con esta encuesta. En este caso, sugiero pasarla al formato virtual, para que sea más fácil consolidar los resultados. **Ideal que esta encuesta sea anónima.**

● Parte 3: Análisis y Planes de Acción

Ahora llegó el momento de analizar los resultados de la encuesta de engagement y definir planes de acción.

3. 1. Supongamos que hoy recibiste los resultados de engagement de tu equipo. Revisa y analiza los resultados de la encuesta, respondiendo las siguientes preguntas: El reporte podrás encontrarlo en la carpeta del Proyecto como Reporte de Engagement Survey.

- ¿Cuál es tu percepción general de los resultados?

Acá debes mencionar los puntos que te llaman la atención en cada categoría del Engagement:

Ej: - Gran parte del equipo no tiene claridad de roles y responsabilidades y tampoco de sus objetivos de corto, mediano y largo plazo.
- 40 % del equipo (approx) no siente que está utilizando sus fortalezas en el trabajo de forma consistente



- ¿Qué pilares del Engagement necesitan atención?

Mis Expectativas / Mis Fortalezas / Mi Desarrollo / Mis Opiniones / Mi Misión y Propósito? Elija los pilares que necesitan planes de acción de mejora.

- ¿Identificas algún problema/solución?

Necesitamos como equipo aclarar roles & responsabilidades y redefinir nuestros objetivos de corto / mediano plazo.

3.2. Defina entre **2 a 4 planes** de acción de engagement para tu equipo.
Utilice el formato SMART: Specific: específico / Measurable: medible / Achievable: alcanzable / Relevant: relevante / Time based: con límite de tiempo.



Pilar de Engagement / Elemento de Bienestar	Resultado esperado	Acción	Responsable por el seguimiento del plan	Métrica	Deadline / Frecuencia
(Ej. Engagement) Mis Expectativas	Que el 100% del equipo tenga su descripción de cargo actualizada.	Cada uno del equipo revisará su descripción de cargos y la actualizará de acuerdo con sus actividades y coordinará una sesión de revisión con el manager	Juan Carlos	Tener 100% de las descripciones de cargo actualizadas	01/XX/XXXX
(Ej. Pilares Bienestar) Bienestar Comunitario y Social	Equipo feliz y conectado, sintiéndose parte de algo mayor.	Participación de la acción de compañía, dónde plantaremos árboles en parques de la ciudad.	Manager	Mínimo 80% del equipo participando del evento.	01/XX/XXXX

3.3. Adicional a los planes de engagement, incluye entre 1 a 3 planes de acción de bienestar que podrían contribuir a tu equipo, considerando los elementos social, financiero, físico y comunitario. Como líder, ¿qué ideas propondrías para mejorar el bienestar de tu equipo?

● Parte 4: Cronograma y plan de seguimiento

4.1. Elaborar un cronograma de seguimiento a los planes de acción individuales y de equipo, incluyendo momentos de celebración de avances, de acuerdo con los planes de acción propuestos en el avance #3. La idea es incluir planes de bienestar integral y no solamente planes de engagement. Puedes utilizar el formato que más te guste y planear las acciones de acuerdo con la realidad de tu empresa.



1) Ejemplo de 3 meses:

JULIO							AGOSTO							SEPTIEMBRE						
D	L	M	M	J	V	S	D	L	M	M	J	V	S	D	L	M	M	J	V	S
					1	2	1	2	3	4	5	6					1	2	3	
3	4	5	6	7	8	9	7	8	9	10	11	12	13	4	5	6	7	8	9	10
10	11	12	13	14	15	16	14	15	16	17	18	19	20	11	12	13	14	15	16	17
17	18	19	20	21	22	23	21	22	23	24	25	26	27	18	19	20	21	22	23	24
24	25	26	27	28	29	30	28	29	30	31				25	26	27	28	29	30	
31																				

2) Incluir el detalle de las actividades de cada mes:

Ej. Agenda agosto:

05/08 - Sesión 30 min con el equipo, para revisar avances del plan de engagement / celebrar éxitos

16/08 & 18/08 – sesiones cortas individuales c/ persona del equipo para hacer seguimiento de sus planes de bienestar individuales

26/08 – Almuerzo de equipo

30/08- Reto Carrera 5K

- ∞ Revisión de planes de acción de engagement (Equipo)
- ∞ Sesión de bienestar (Individual - c/ cada integrante del equipo)
- ∞ Actividad de bienestar – trabajo voluntario, ejercicio físico, clase de cocina saludable, etc. (Equipo)
- ∞ Fun time – Salidas de equipo, almuerzos, HHs, Off-sites, etc (Team)

4.2. Haz un listado de acciones que vas a discutir con tus pares o con el equipo de Recursos Humanos, con las iniciativas que serían de gran impacto para toda la organización, más allá de tu equipo. Hazlo en orden de prioridad y en caso que exista la posibilidad de ejecutar alguna de esas acciones, ponla en tu cronograma.



Considera todos los elementos del bienestar, salud mental, liderazgo, felicidad, manejo de tiempo y todo lo que hemos visto durante el curso / Escoge acciones que puedan favorecer más de un elemento a la vez.

Ejemplos de acciones:

- Crear un grupo de embajadores de bienestar en las diferentes áreas de la organización
- Incluir conversaciones de bienestar en las discusiones formales de desempeño de la empresa.

crehana 