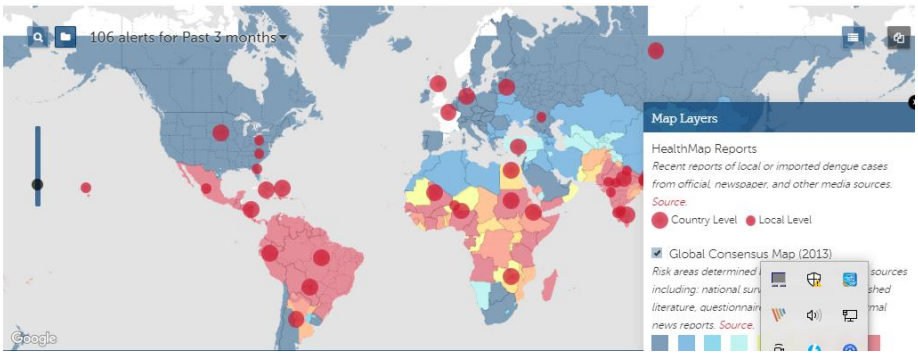
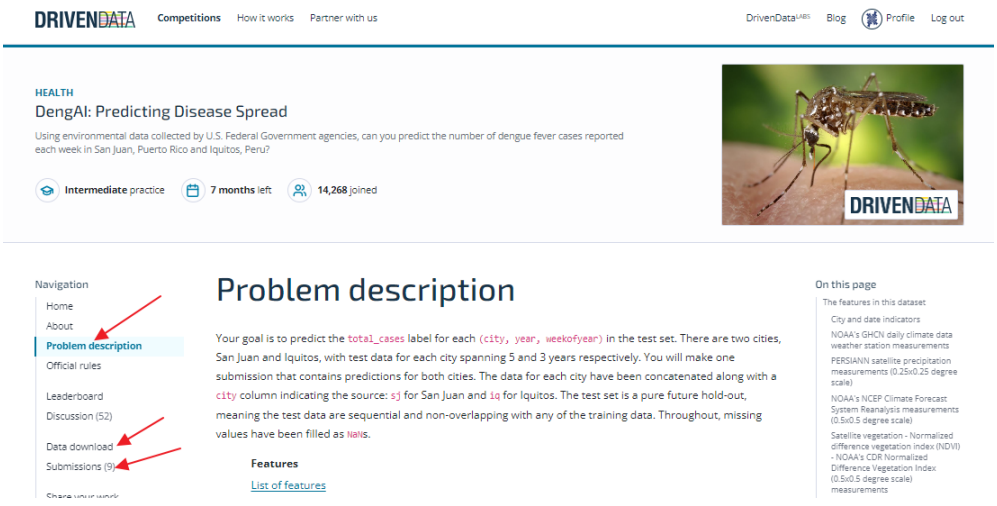
Actividad 3.6 - DengAI: predicción de la propagación de enfermedades



El objeto de esta actividad es participar en la competición de ofrecida de la web de DrivenData denominada: DengAI: Predicting Disease Spread.

Para ello accederemos a la siguiente web y nos crearemos un usuario:

Título: DengAI: Predicting Disease Spread
Url: https://www.drivendata.org/competitions/44/dengai-predicting-disease-spread/



La actividad consiste en subir a dicha web un fichero csv con la estimación que hayamos obtenido al aplicar los modelos que consideres oportunos. Podrás observar que se pueden realizar hasta un máximo de 3 subidas diarias y la propia web realizará una valoración de tu solución. (La valoración se realiza utilizando el MAE como criterio de valoración de calidad)

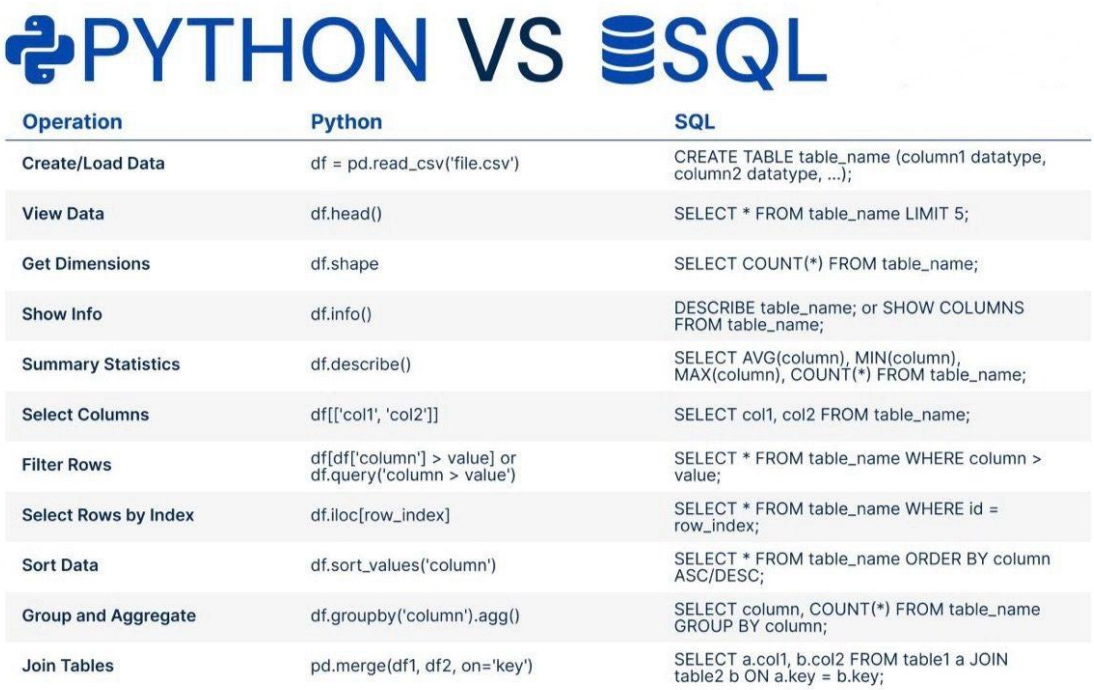
Hay que realizar todo el proceso completo de importación del dataset, ajuste de características, selección de características, entrenamiento y selección de un modelo para entrenarlo, predicción y subida del fichero para que la web lo valores y clasifique en la competición.

Consideraciones a tener en cuenta:

- Es necesario realizar pruebas con al menos tres modelos diferentes considerando que han de utilizarse como mínimo **NaiveBayes, KNN** y **otro modelo** que consideres oportuno
- Realizar pruebas de hiperparametrización con las dos técnicas explicadas: GridSearch y Random Search.

Una vez hayas realizado las diferentes pruebas, has de generar un documento pdf donde:

Expliques con detalle tu solución. Ha de contener capturas de tu posicionamiento en la competición con la clasificación que has obtenido en la competición con los diferentes SUBMITs realizados. Sería muy interesante que expliques las mejoras o no, que has obtenido con las diferentes pruebas.



Operation	Python	SQL
Create/Load Data	df = pd.read_csv('file.csv')	CREATE TABLE table_name (column1 datatype, column2 datatype, ...);
View Data	df.head()	SELECT * FROM table_name LIMIT 5;
Get Dimensions	df.shape	SELECT COUNT(*) FROM table_name;
Show Info	df.info()	DESCRIBE table_name; or SHOW COLUMNS FROM table_name;
Summary Statistics	df.describe()	SELECT AVG(column), MIN(column), MAX(column), COUNT(*) FROM table_name;
Select Columns	df[['col1', 'col2']]	SELECT col1, col2 FROM table_name;
Filter Rows	df[df['column'] > value] or df.query('column > value')	SELECT * FROM table_name WHERE column > value;
Select Rows by Index	df.iloc[row_index]	SELECT * FROM table_name WHERE id = row_index;
Sort Data	df.sort_values('column')	SELECT * FROM table_name ORDER BY column ASC/DESC;
Group and Aggregate	df.groupby('column').agg()	SELECT column, COUNT(*) FROM table_name GROUP BY column;
Join Tables	pd.merge(df1, df2, on='key')	SELECT a.col1, b.col2 FROM table1 a JOIN table2 b ON a.key = b.key;

Fuente:
https://www.linkedin.com/feed/update/urn:li:activity:7290306350243274752/?utm_source=share&utm_medium=member_desktop

Título: Modelos de machine learning: Guía básica para principiantes
Url: https://planetachatbot.com/modelos-de-machine-learning-guia-basica-para-principiantes/

	Criterios para valorar Proyecto/Reto	
	Peso %	**Tareas (Se evalúa desde 0 hasta 10)**
	5	Utilizar el drive o GitHub como origen de ficheros para la importación del dataset.
	5	Importación del dataset: Preparación de los datos: Normaliza, ajusta la calidad de los datos.
	5	Selección de caracteristicas: Utiliza métodos no gráficos para la selección de características.
	5	Selección de caracteristicas: Utiliza herramientas gráficas para la elección de las características.
	10	Además de la división de los datos de train y test, incorpora la utilización de datos de validación.
	10	Entrenamiento: Modelo 1º NaibeBayes - Desarrolla las diversas pruebas propuestas para la selección y justificia el criterio de calidad para la selección del modelo. Utiliza Cross Validation
	10	Entrenamiento: Modelo 2º KNN - Desarrolla las diversas pruebas propuestas para la selección y justificia el criterio de calidad para la selección del modelo. Utiliza Cross Validation
	10	Entrenamiento: Modelo 3º de elección libre - Desarrolla las diversas pruebas propuestas para la selección y justificia el criterio de calidad para la selección del modelo. Utiliza Cross Validation
	10	Entrenamiento - Uso de gráficos: Integra el uso de gráficos para obtener comparativas en el entrenamiento de los modelos.
	10	Predicción: Utiliza herramientas gráficas para ayudar a entender la precisión de los resultados obtenidos.
	5	Predicción: Describe con claridad una valoración de los resultados obtenidos.
	5	Submit del fichero con la predicción y captura de la valoración /posicionamiento obtenido en la competición.
	5	Propone soluciones creativas e innovadoras.
	5	El pdf final tiene una portada., utiliza un índice, apartado de conclusiones y referencias (web). Se hace mención a referencias externas, no recogidas en el material suministrado.
Total	**100**	

Formato de entrega

Es **obligado** entregar un fichero en un Archivo PDF con capturas del código y resultados obtenidos, así como la url de GitHub y Google Colab donde has publicado el código.

- Nombrar el archivo siguiendo el siguiente patrón:

  SNS_ACT3_6_NombreApellidos.pdf