

Questions and answers

1. Option C:

1.1. Was it possible to demonstrate the hypothesis? Why?

Como ya expliqué en el otro archivo, la hipótesis 1 se pudo demostrar parcialmente para los estadísticos de dispersión en horario diurno, mientras, para el resto de situaciones no se satisfizo dicha hipótesis.

La hipótesis 2 se demostró en cada resultado que se iba obteniendo, siendo esta hipótesis, todo un éxito.

1.2. What can you conclude about your data study?

Puedo concluir que los datos estaban muy bien estructurados, incluso antes de haberlos trabajado. Esta estructura hizo que me fuese sencillo trabajar con ellos.

1.3. What would you change if you need to do another EDA project?

Cambiaría, quizás, el enfoque que le he dado, pues este ha sido más como si de un artículo científico se tratase en lugar de ser un enfoque puramente pedagógico en el cual obtener mucho más resultados aún si no fuesen necesarios para demostrar la hipótesis.

1.4. What do you learn doing this project?

Respecto a data science he mejorado mi conocimiento de las librerías pandas y matplotlib, así como una introducción a basemaps.

He aprendido como organizar proyectos de data science y a trabajar con módulos distintos.

Fuera del tema computacional, he aprendido mucho sobre geografía y climatología.

2. Option B:

2.1. Which are the columns with the highest correlation?

Pregunta compleja debido a la cantidad de dataframes que he usado, sin embargo, desde el punto de vista del dataframe global, las columnas con mayor correlación fueron la media y la mediana.

En el dataframe de meses, cada mes con su contiguo tenía una correlación de 0.9 aprox.

3. Option A:

3.1. Are there outliers or some rare data?

Hubo mucho valor NaN, pero si se pregunta por valores anómalos, no hubo muchos dado que el método de recogida de datos de la NASA es muy eficiente y no produce datos anómalos. Quizás la NASA limpie los datos antes de ofrecerlos, pero si no es así, no hay prácticamente datos anómalos.

3.2. What are the columns that have more repeated values?

Más que las columnas, la clave está en entender cuáles son las filas (latitudes) que poseen más valores repetidos. En este caso dichas son las cercanas al Ecuador, 0 grados.