

- Brief description of the data set and a summary of its attributes:
 - The dataset is on the topic of Power Consumption of Tetouan City and is a UCI dataset with six features and three targets. One feature is the date of the recording over the month of January 2017 for the city, with the recording being added for every ten-minute period of the day. The temperature, humidity, wind speed, general diffuse flows, and diffuse flows for the city are recorded. Finally, the zone 1, 2, and 3 power consumptions of the city are recorded as targets.
- Initial plan for data exploration:
 - The plan is to collect quantile data and summary statistics from the dataset, along with heatmaps for the data correlations, and graphs to represent the relationship between temperature and humidity, general diffuse flow and diffuse flow, diffuse flow and power consumption, and power consumptions for all zones.
- Actions taken for data cleaning and feature engineering:
 - The data is checked for null entries, and data during the day of January second is used to checking for correlations and scatter plot building, and data from January second and third between 00:00 and 23:50 are used for hypothesis testing for diffuse flows.
- Key Findings and Insights, which synthesizes the results of Exploratory Data Analysis in an insightful and actionable manner:
 - Power consumption across zones is strongly correlated with one another.
 - Power consumption across all zones increases sharply as temperature increases past 12.50 degrees and decreases sharply after temperature drops below 15 degrees.
- Formulating at least 3 hypothesis about this data:
 - Humidity is roughly the same across various days of the week.
 - $H_0 : u_1 \neq u_2$
 - $H_1 : u_1 = u_2$
 - Zone one power consumption is the same across days of the week
 - $H_0 : u_1 \neq u_2$
 - $H_1 : u_1 = u_2$
 - Higher temperature or humidity are associated with higher power consumption.
 - $H_0 : u_1 = u_2$
 - $H_1 : u_1 > u_2$
- Conducting a formal significance test for one of the hypotheses and discuss the results:
 - According to a t-test of the data, we notice that between the hours of 00:00 and 23:50 during the days of 02 and 07 of January 2017, the p-value of zone one power consumption across both days is 0.031031563896130217, which is less than an alpha of 0.05, meaning that we reject the null hypothesis that there is no difference between power consumption levels in Zone 1 across both days.
- Suggestions for next steps in analyzing this data:
 - Create a loop to compare the t-test results for humidity values across days of the week.
 - Use PCA to combine temperature, humidity, and wind speed features and test correlation against power consumption levels.
- A paragraph that summarizes the quality of this data set and a request for additional data if needed:

- The dataset has a high quality and contains no null values or missing pieces of information. Extra data for months other than January could help with discovering power consumption changes across months of the year or seasons. The data contained some outliers – as was indicated by the data having z-scores with a greater magnitude than 3 – which include 23 outliers related to general diffuse flows, 1361 outliers related to diffuse flows, 88 outliers related to temperature, 656 outliers related to zone 3 power consumption, and 1 outlier related to zone 2 power consumption.