**EECS 731 - Introduction to Data Science**

**Topic: Jimmy Wrangler, Data Explorer**

* Repo name: Explore\_Data

The project structure for this Data Science assignment is set up in the mentioned Repository under my GitHub: <https://github.com/Priyanka1527/Explore_Data>

* Required frameworks installed : Python Jupyter Notebook, Plotly
* Industry Chosen: Ride Sharing Industry

Data Source: <https://www.kaggle.com/datasf/san-francisco>

* Public Datasets from this industry: **bike\_station, bike\_trips** (I had to truncate and select few Rows on this dataset as it was huge) committed under Explore\_Data/data
* Datasets were loaded into Pandas dataframe and transformed to produce a single dataset

**result\_data.ipynb** committed under Explore\_Data/notebooks

* **Ideas**: On combining the 2 datasets I could come up with the count of how many trips were taken near to a particular landmark in the area. This could help the Rideshare company to have more stations, docs and bikes installed in that particular area due to the increasing demand.
* References:

<https://plot.ly/python/aggregations/>

<https://plot.ly/python/ipython-notebook-tutorial/>