

**Rohan Sanjay Kore**

**x19214413@student.ncirl.ie**

### **PERSONAL PROJECT JOURNAL**

#### TASK 1: Project Topic Selection -

- Held the group meet on teams and discussed various domain which can be selected and will be appropriate for the project. Everyone suggested topics from domains like sports, healthcare, Sales. I found out an interesting domain and related datasets from NYC open directory and discussed it with members. Further after discussion with professor we finalized the topic factors impacting NYC housing.
- Each team member equally contributed their ideas in the Teams group session.

#### TASK 2: Data Acquisition –

- Each team member went through several online data repositories for procuring the data as per discussed domain.
- Using NYC Open Data website took NYC property sales data using SODA API.

##### **Challenges encountered :-**

- To find a common link to answer our research questions was important.
- After fetching data from soda API it was giving only 1000 rows. So had to increase it to 1000000 rows for better visualisations and understanding the data.

#### TASK 3: MongoDB and Preprocessing –

- Further the NYC sales dataset was taken by me from NYC Open Data website. I established the connection with mongoDB Atlas to push json file into mongoDB for structured data and later Data Transformation and Preprocessing steps were carried out.
- Removing NaN values, dropping unrelated columns, breaking down columns like date into month and year and changing data types

#### TASK 4: PostgreSQL and Visualization -

- In this, first the connection was established, and database was created. Later Table was created and my record was pushed into this table.
- The loaded data was then read and stored into DataFrame.
- After getting our data into Dataframe, I visualized my data to get meaningful insights.

#### TASK 5: Reports-

- Contributed to group report and video presentation.