NYC Traffic Data Violation Analysis

1. Connect to the API:
Below is the python code used to pull recent 10,000 rows from the Data Source: https://data.cityofnewyork.us/City-Government/Open-Parking-and-Camera-Violations/nc67-uf89/about_data
And csv file was downloaded to local.

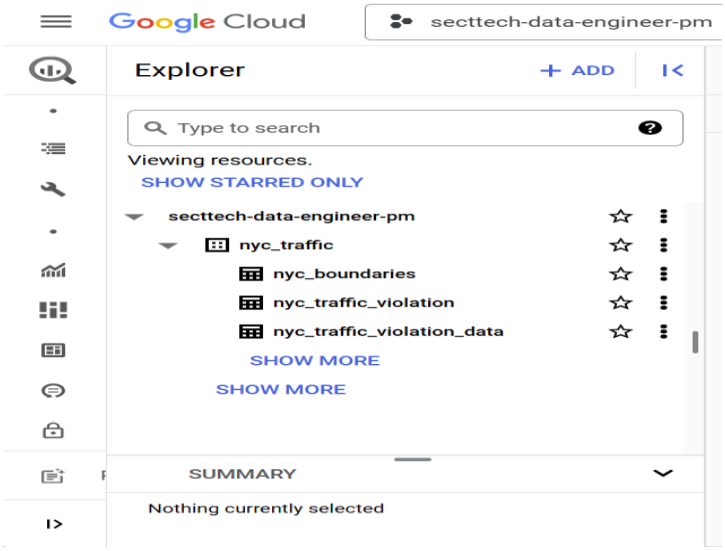
Snippet of Python code

```python
import pandas as pd
import requests

# Define the URL for the NYC Open Data API
dataset_url = "https://data.cityofnewyork.us/resource/nc67-uf89.json"

# Define parameters for sorting and selecting recent 10000 records
limit = 10000  # Number of records to fetch
params = {'$limit': limit, '$order': 'issue_date DESC'}

try:
    # Fetch the data with sorting and selecting top records
    response = requests.get(dataset_url, params=params)
    # Check if the request was successful
    response.raise_for_status()
    # Attempt to parse the JSON data
    data = response.json()
    # Load data into a pandas DataFrame
    df = pd.DataFrame(data)
    # Display or save the result
    print(df)
        # Specify the directory where you want to save the file
    save_directory = 'C:/documents/'
        # Save the file with the specified directory
    df.to_csv(f'{save_directory}recent_10000_records.csv', index=False)

except requests.exceptions.RequestException as e:
    print(f"Request error: {e}")
```

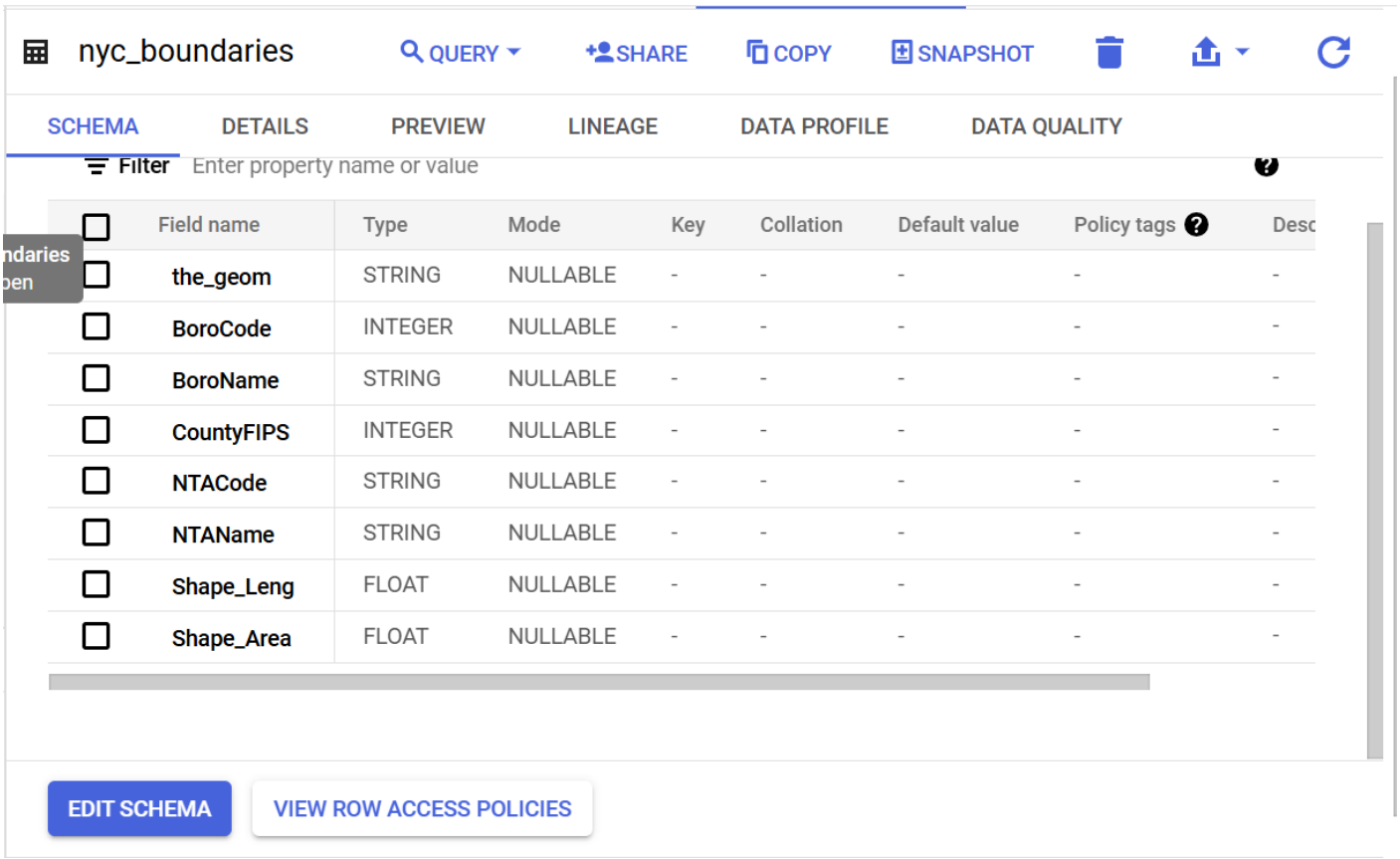
2.Load data into BigQuery:

Created dataset inside given project with name `nyc_traffic` and created table `nyc_traffic_violation` from input csv file using bigquery upload option.
Created a lookup table inside this dataset namely `nyc_boundaries` as external source to get region's data.

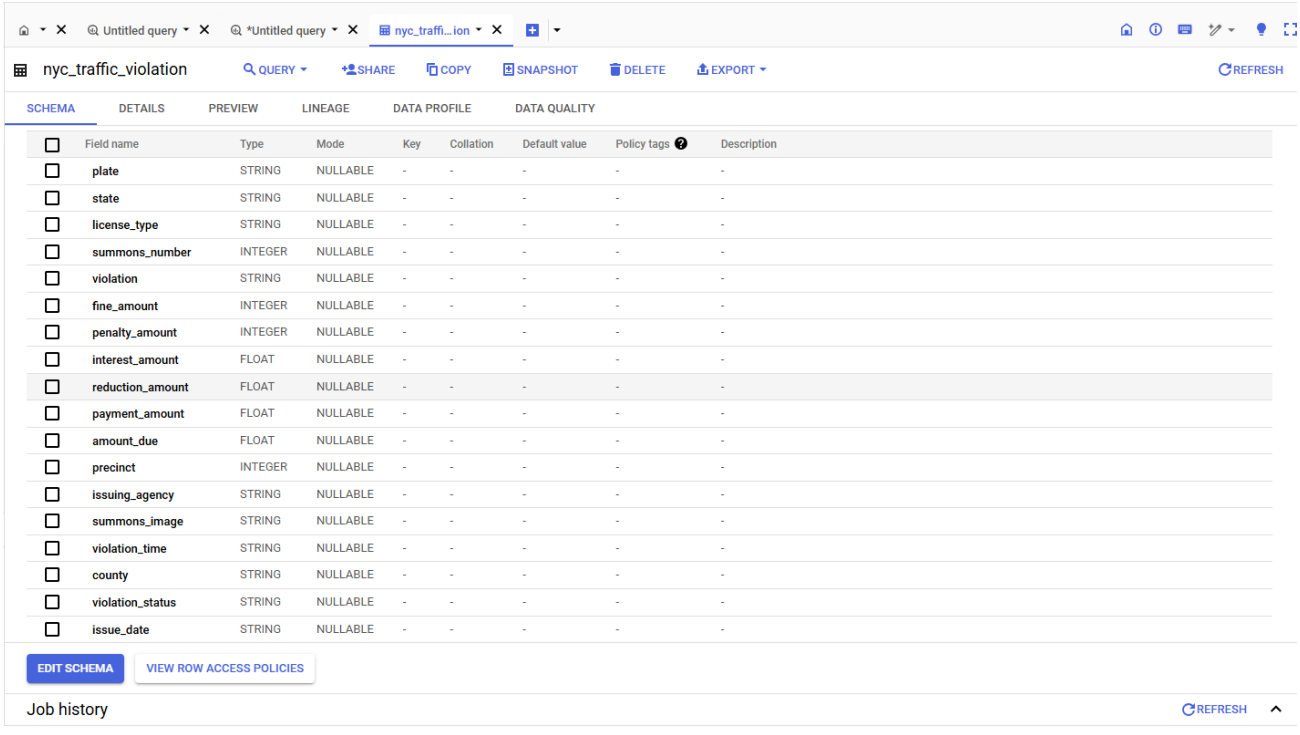
Snippet of BigQuery table and dataset



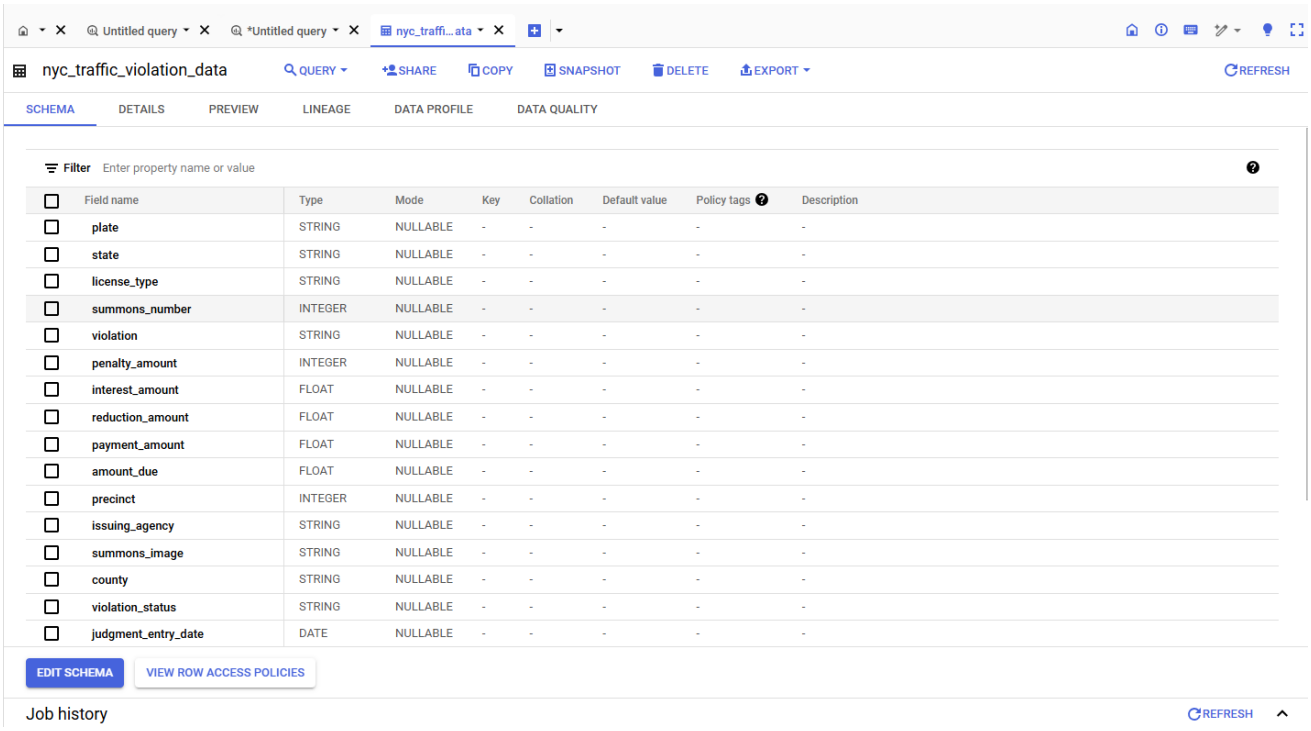
Snippet of Table Structure nyc_boundaries



Snippet of Table Structure for nyc_traffic_violation



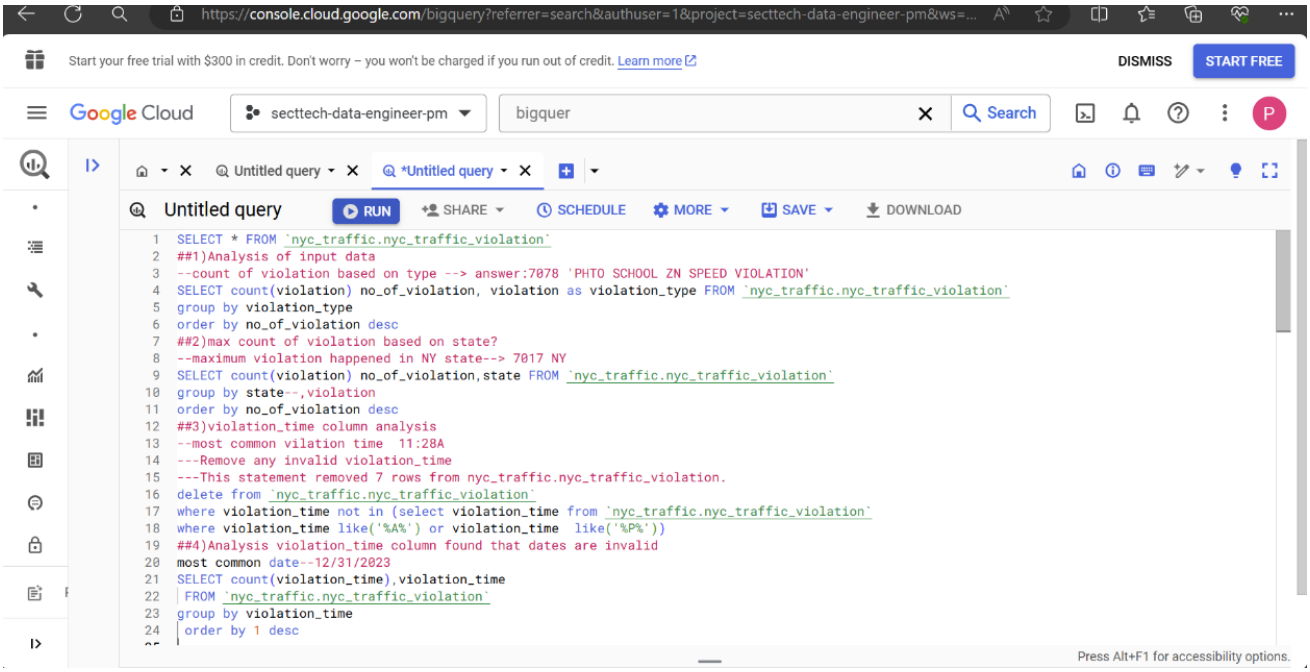
Snippet of Table Structure for nyc_traffic_violation_data



3. Data Transformations:

Below are code snippets used to Analyse the data and found invalid entries in violation time column and time format was also not proper and while importing data fine_amount column was created as integer but actually should been integer.

Snippet of Data Analysis SQL Queries

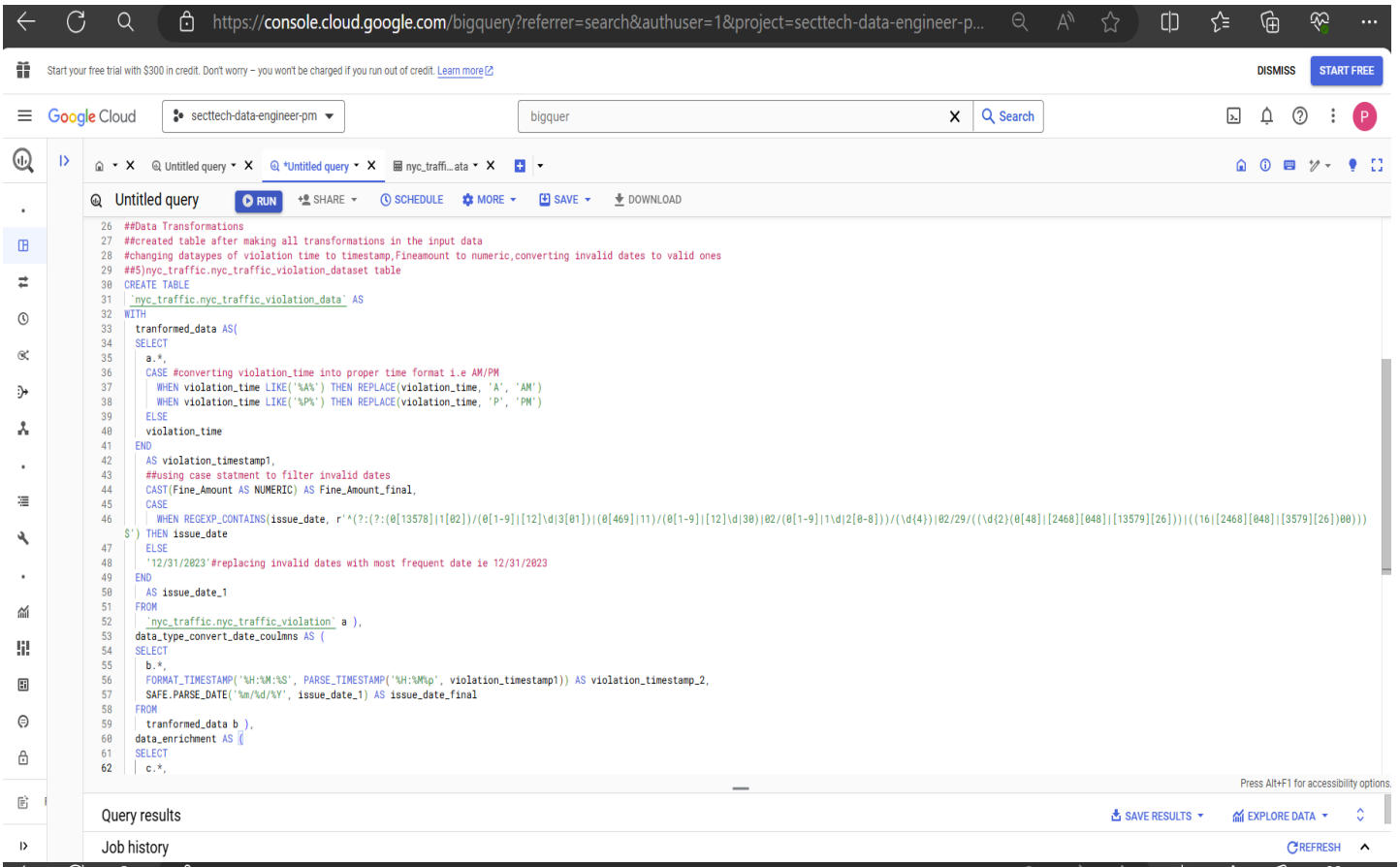


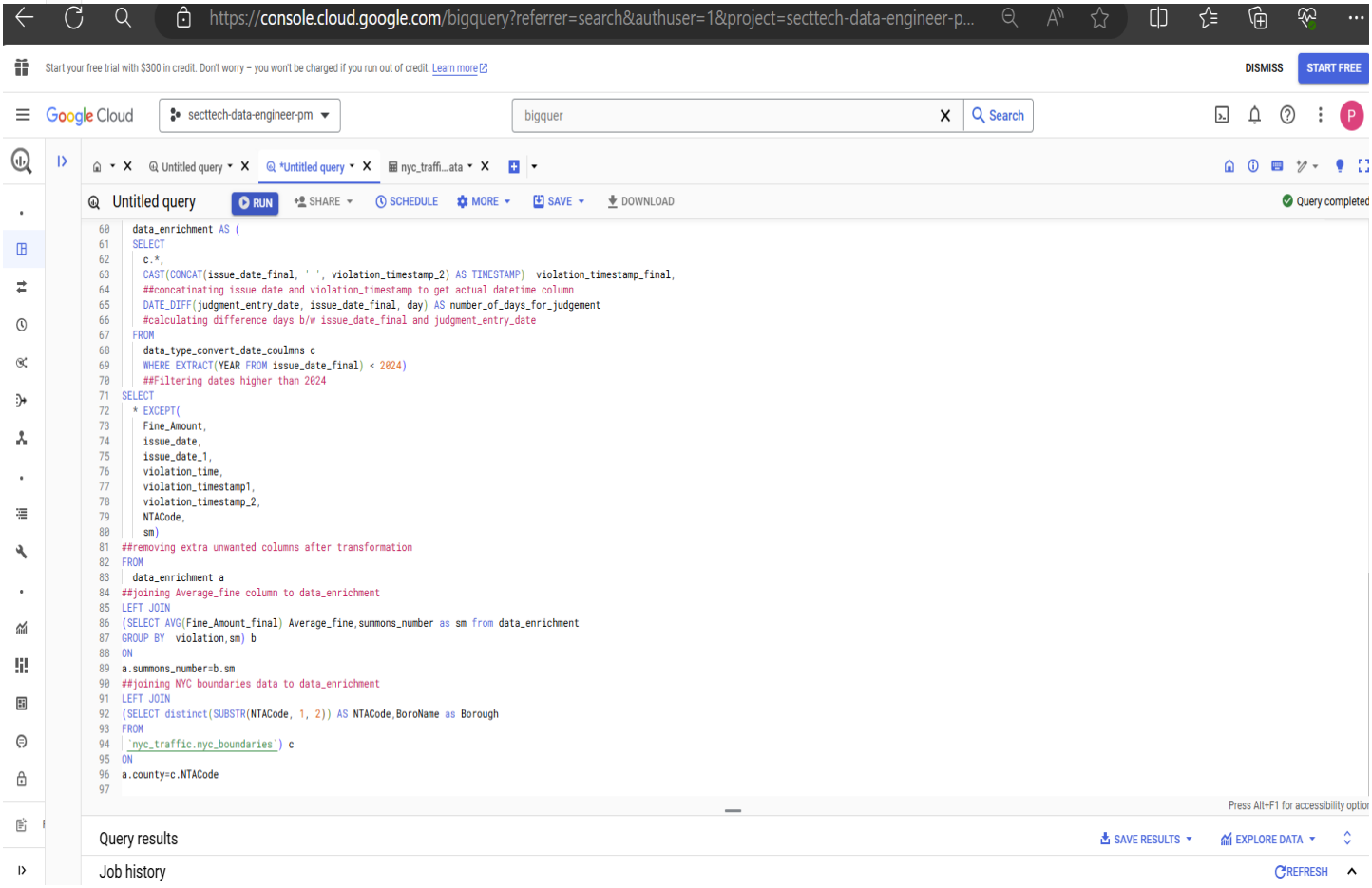
Below transformation have been applied to the input data

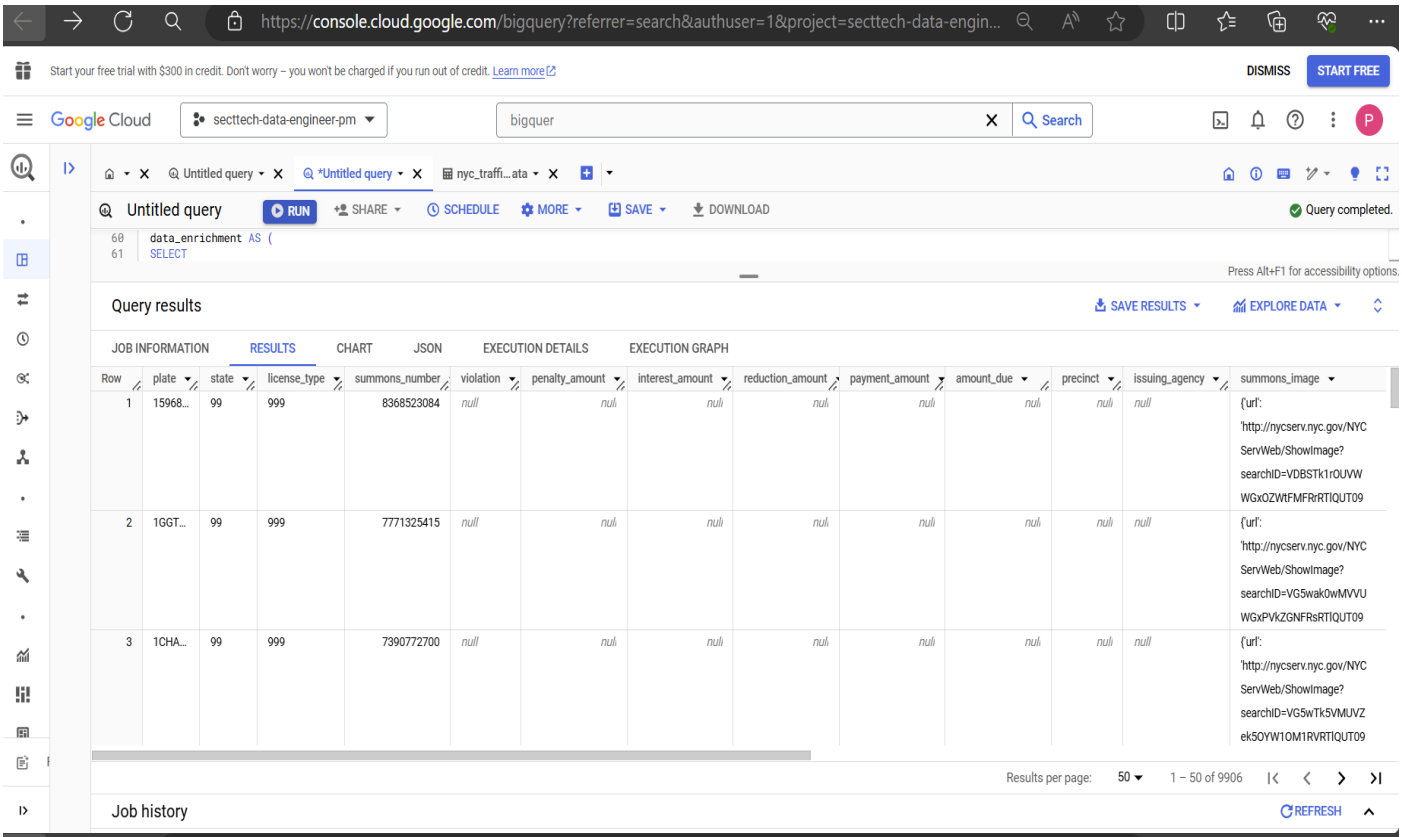
- Column violation_time records were first converted to standard timings by using CASE statement and replace function.
- Column fine_amount data_type was changed to numeric as it cannot be negative (integer type) and renamed as fine_amount_final .
- REGEXP_CONTAINS was used to identify faulty/invalid issue date and replaced with most frequent appearing date.
- Concatenation of issue_date and violation_time to get Violation timestamp column,

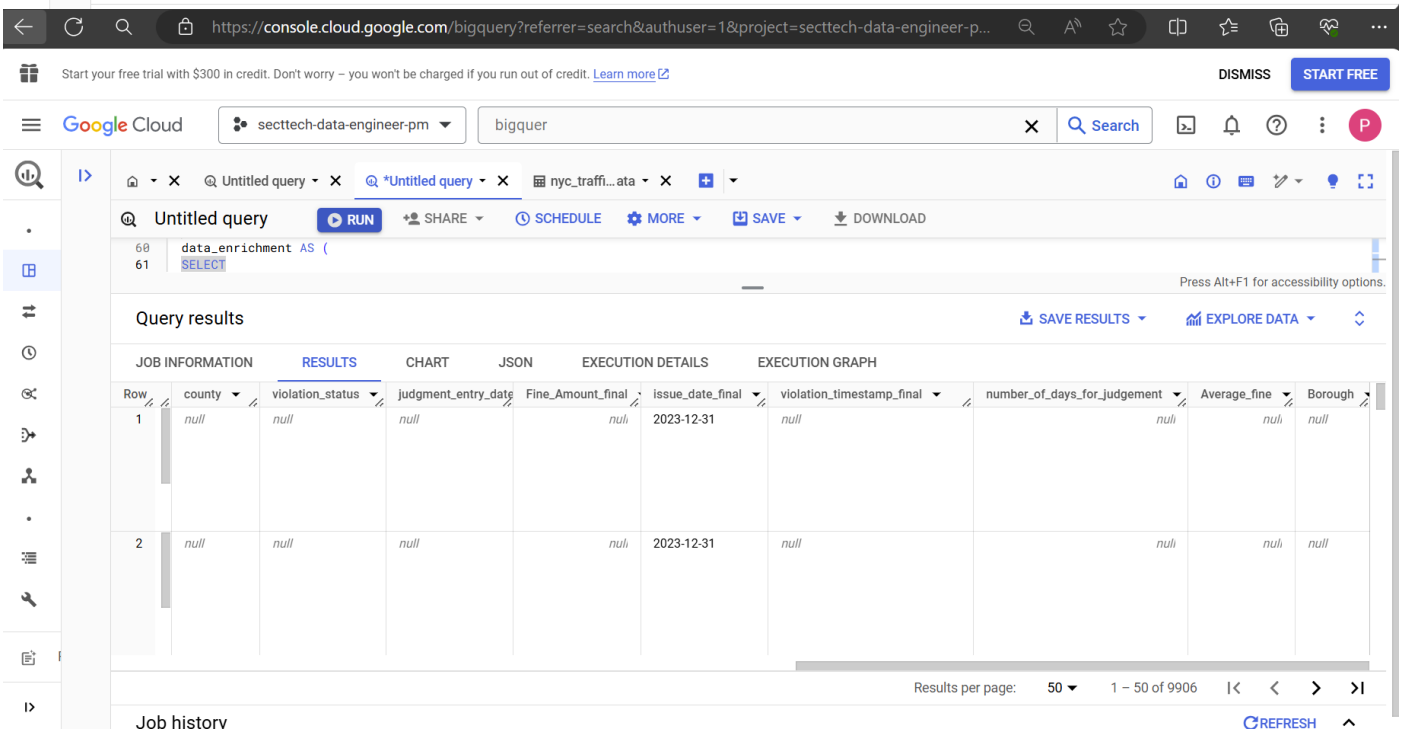
- Additional columns like number_of_days_for_judgement,Average_fine ,borough(was obtaining by joining nyc_boundaries data using NTA code to county)

Snippet of Transformed Data Creation






https://console.cloud.google.com/bigquery?referrer=search&authuser=1&project=secttech-data-engin...
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
START FREE
Google Cloud
secttech-data-engineer-pm
bigquer
Search
Untitled query
*Untitled query
nyc_traffi...ata
RUN
SHARE
SCHEDULE
MORE
SAVE
DOWNLOAD
Query completed.
60 data_enrichment AS (
61 SELECT
Press Alt+F1 for accessibility options.
Query results
SAVE RESULTS
EXPLORE DATA
JOB INFORMATION
RESULTS
CHART
JSON
EXECUTION DETAILS
EXECUTION GRAPH
Row plate state license_type summons_number violation penalty_amount interest_amount reduction_amount payment_amount amount_due precinct issuing_agency summons_image
1 15968... 99 999 8368523084 null null null null null null null null {'url': 'http://nycserv.nyc.gov/NYCServWeb/ShowImage?searchID=VDBSTk1rOUVWWGxOZWtFMFRrRTlQUT09
2 1GGT... 99 999 7771325415 null null null null null null null null {'url': 'http://nycserv.nyc.gov/NYCServWeb/ShowImage?searchID=VG5wak0wMVVUWGxPVkZGNFRsRTlQUT09
3 1CHA... 99 999 7390772700 null null null null null null null null {'url': 'http://nycserv.nyc.gov/NYCServWeb/ShowImage?searchID=VG5wTk5VMUVZek5OYW1OM1RVRTlQUT09
Results per page: 50
1 – 50 of 9906
Job history
REFRESH

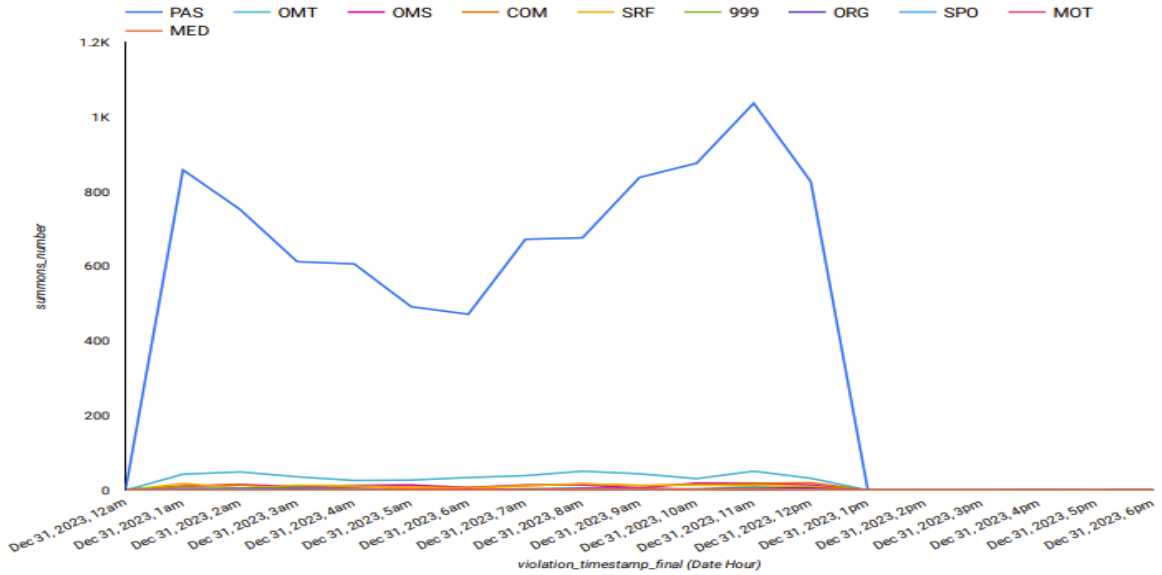

https://console.cloud.google.com/bigquery?referrer=search&authuser=1&project=secttech-data-engineer-p...
Start your free trial with $300 in credit. Don't worry – you won't be charged if you run out of credit. Learn more
DISMISS
START FREE
Google Cloud
secttech-data-engineer-pm
bigquer
Search
Untitled query
*Untitled query
nyc_traffi...ata
RUN
SHARE
SCHEDULE
MORE
SAVE
DOWNLOAD
60 data_enrichment AS (
61 SELECT
Press Alt+F1 for accessibility options.
Query results
SAVE RESULTS
EXPLORE DATA
JOB INFORMATION
RESULTS
CHART
JSON
EXECUTION DETAILS
EXECUTION GRAPH
Row county violation_status judgment_entry_date Fine_Amount_final issue_date_final violation_timestamp_final number_of_days_for_judgement Average_fine Borough
1 null null null null 2023-12-31 null null null null
2 null null null null 2023-12-31 null null null null
Results per page: 50
1 – 50 of 9906
Job history
REFRESH

4.Data Visualization:
Looker Studio has been used to visualize and explore insights from transformed data.

Snippet of Visualization done in Looker Studio

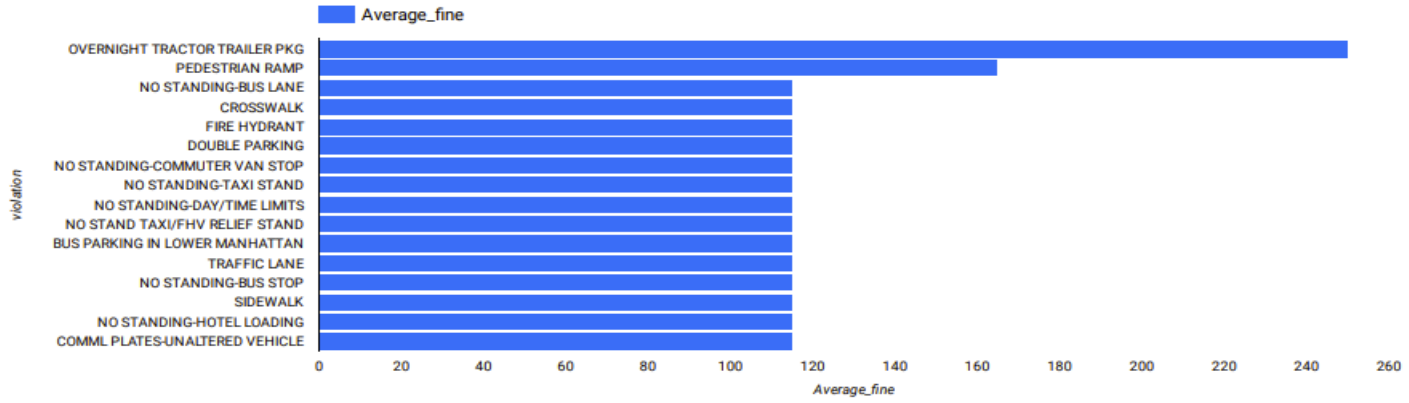
NYC Traffic Violation Analysis

Time series Analysis of violation by License type

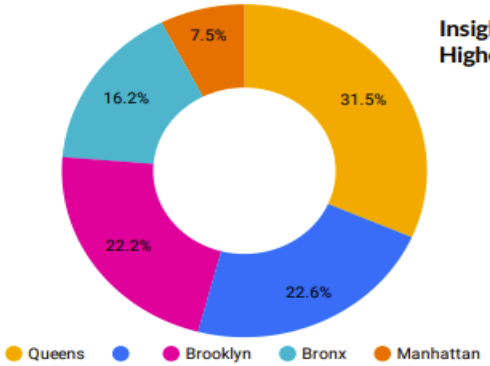


Insights-
1) Most of the violations happened between the time 12AM TO 1PM
2) Highest number of violations are made by license type PAS

Average fine violation wise

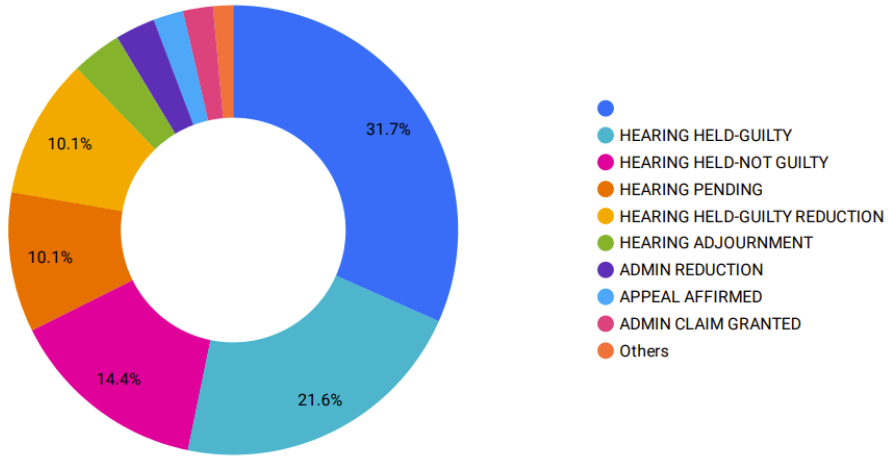


Average fine violation borough wise



Insights-
Highest number of violations are in Queens region

Violation Status



Insights-
18% of violations has violation status as Hearing Held-not Guilty