

ALBANY CITY OUTSTANDING PARKING TICKETS

IST 722: PROJECT REPORT

Group 3:

- Vignesh Chandramouli
- Anoop S. Daniel
- Vimal Raj Susai Raj
- Nandakumar Udayakumar
- Hariharan Ramanarayanan

Guided By: Prof. Jonathan Fox

TABLE OF CONTENTS

DESIGN DOCUMENT:.....	2
1. PROJECT OVERVIEW	2
1.1. Project Scope	2
1.2. Team Members and Roles.....	2
2. PROJECT ANALYSIS	3
2.1. Business Process	3
3. DIMENSIONAL DESIGN	4
3.1. Fact/Derived Fact Worksheet	4
3.2. Detailed Dimensional Design Worksheet	4
3.3. Dimensional Hierarchies	5
3.4. Snowflake Model Diagrams.....	6
3.5. ETL Specifications	6
IMPLEMENTATION:.....	7
4. ROLAP SCHEMA IN SQL SERVER:.....	7
STAR SCHEMA DIAGRAMS	9
5. SSIS ETL CODE/PACKAGES.....	9
6. Analysis Services MOLAP Database:.....	12
ANALYSIS AND REPORTING.....	13
7. Front-End Linkage.....	13
8. Observation and Analysis:	15
ANNEXURES:	20

DESIGN DOCUMENT:

1. PROJECT OVERVIEW

The focus of this endeavor is to showcase our learning outcomes from the course and to demonstrate our ability to work on a data warehouse project dealing with all aspects of data design, implementation and analysis. This project would be an opportunity for the student group to demonstrate our progress in describing the various database constructs, components of the data warehouse, designing and implementation of the data warehouse and BI components using our understanding of SQL concepts as well as dimensional modelling techniques.

Business Context:

The City of Albany is facing a rise in the number of outstanding unpaid parking tickets and traffic violations, and the penalties that are imposed as a result are relatively high. The open data team of the city of Albany posted a data set containing unpaid parking violations from January 2010 through November 17, 2015 (*Annexures*). The city of Albany would benefit from any form of actionable insights derived from the data that might help them analyze and comprehend the root cause of the problem first to try and take preventive measure. It would help to compare the performance of different population demographic in terms of the number of violations and the total dollar amount in terms of original fine amounts and penalties incurred as a result of nonpayment.

The data warehouse team would look to compare if the citations are frequent in in-state (State of New York) or out of state, in which age groups it is predominant and categorize them based on geographies to observe if any specific patterns emerge that would show causation and correlation.

1.1. Project Scope

The scope of the project is to design and implement a data warehouse with star schema dimensional modelling technique, identify dimensional hierarchies, establish ETL specifications such as source to target mapping, implement ROLAP schema in SQL server, SSIS ETL code packages, MOLAP cube to match the star schema and finally perform the analysis based on retrieving the stored data using BI tools. This is done by means of creating interactive dashboards and generating business intelligence reports that produce actionable insights to help understand the root cause of the problem.

1.2. Team Members and Roles

Group 3: Vignesh Chandramouli,
 Anoop S. Daniel,
 Vimal Raj Susai Raj,
 Nandakumar Udayakumar,
 Hariharan Ariyanayakipuram Ramanarayanan

Business Lead(s): Group member(s) that are primarily responsible for the project initiative. As this final project is a group project, the entire group and its members take accountability and responsibility for the completion and overall success of this project.

- Vimal Raj Susai Raj
- Nandakumar Udayakumar

Project Manager(s): Primarily responsible for managing the project, from project plan till implementation.

- Vignesh Chandramouli
- Anoop S. Daniel

Business Analyst: Gathers the various requirements that is necessary for the overall completion of this project.

- Hariharan Ariyanayakipuram Ramanarayanan
- Vimal Raj Susai Raj

Data Architect: Group member(s) primarily responsible for the design and implementation for dimensional modeling.

- Nandakumar Udayakumar
- Vignesh Chandramouli

ETL Architect: Oversees the ETL design and implementation

- Anoop S. Daniel
- Hariharan Ariyanayakipuram Ramanarayanan

BI Architect: Oversees the Business Intelligence design and implementation

- Nandakumar Udayakumar
- Vimal Raj Susai Raj

2. PROJECT ANALYSIS

2.1. Business Process

Outstanding parking tickets reporting: Tracking the number of parking citations as well as total dollar amounts.

- i. Business Process: Outstanding tickets
- ii. Fact: How many (most)
- iii. Product Type: Parking
- iv. Dimension: Vehicles, States and, Owners

3. DIMENSIONAL DESIGN

3.1. Fact/Derived Fact Worksheet

Title	AccountKey	Description
Citation_Fact	license_key	Foreign Key. Uniquely identifies each license.
	state_key	Foreign Key. Uniquely identifies each states.
	owner_key	Foreign Key. Uniquely identifies each owner.
	citation_key_no	Primary Key. Uniquely identifies each citation.
	original_Amount	Gives the original amount of the ticket.
	penalties	Provides details about the penalties for each citation corresponding to the owner.
	balance_due	Indicates the balance due or outstanding

3.2. Detailed Dimensional Design Worksheet

Title	AccountKey	Description
License_Dim	License_key	Primary Key. Uniquely defines each license.
	Vehicle_license_plate	Business Key. Indicates the unique license plate for vehicles.
	Car_type	Gives the type of car
Owner_Dim	Owner_key	Primary Key. Uniquely identifies each owner.
	Driver_id	Business Key. Uniquely provides ID to each driver.
	First_Name	Provides first name of the owner
	Last_Name	Provides last name of the owner
	Age	Gives the age of the driver
	Gender	Indicates the gender of the owner

State_Dim	State_key	Primary Key. It uniquely provides ID to each state.
	Vehicle_state	Business Key. Provides unique ID to each state vehicle.
	stateDesc	Provides description about each state
	Vehicle_region	Gives information about the region of the vehicle.

3.3. Dimensional Hierarches

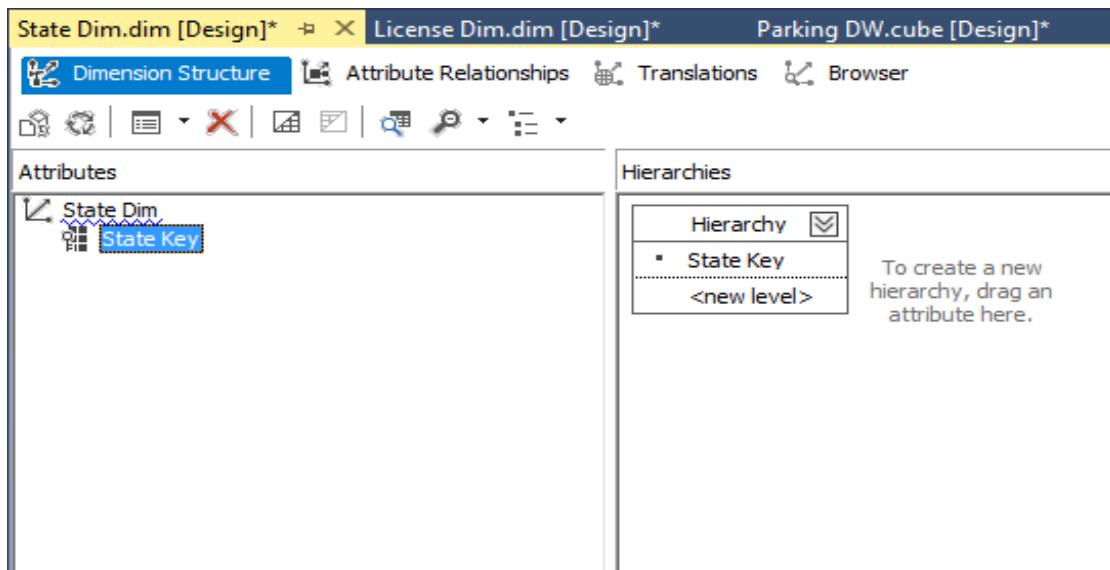
The screenshot shows two separate instances of the Microsoft Analysis Services (SSAS) Dimension Structure tool.

Top Window (License Dim.dim [Design]):

- Attributes:** Shows the dimension "License Dim" with its attribute "License Key".
- Hierarchies:** A hierarchy is being created for the "License Key" attribute. The hierarchy tree shows "License Key" at the top level, with a placeholder "<new level>" below it. A tooltip indicates: "To create a new hierarchy, drag an attribute here."

Bottom Window (Owner Dim.dim [Design]):

- Attributes:** Shows the dimension "Owner Dim" with its attribute "Owner Key".
- Hierarchies:** A hierarchy is being created for the "Owner Key" attribute. The hierarchy tree shows "Owner Key" at the top level, with a placeholder "<new level>" below it. A tooltip indicates: "To create a new hierarchy, drag an attribute here."

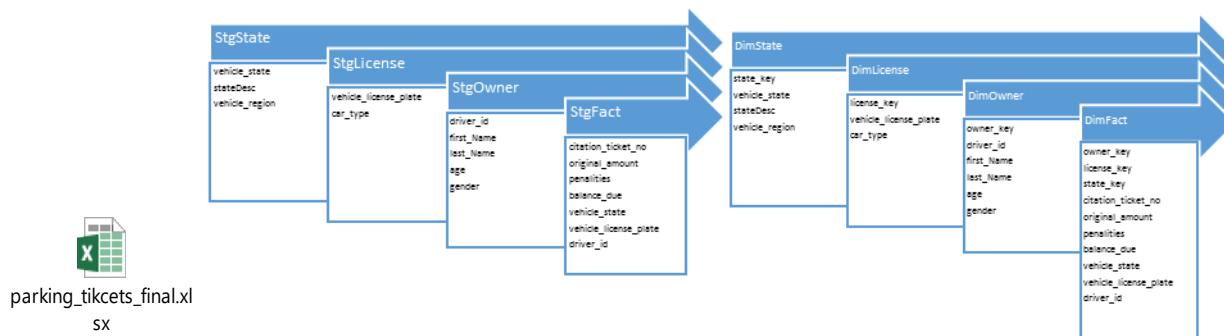


3.4. Snowflake Model Diagrams

This project deals with the outstanding ticket count of parking citations in Albany City. It does not have any relationships that require the need for snow-flaking and hence, the snow-flaking model does not apply in this scenario.

3.5. ETL Specifications

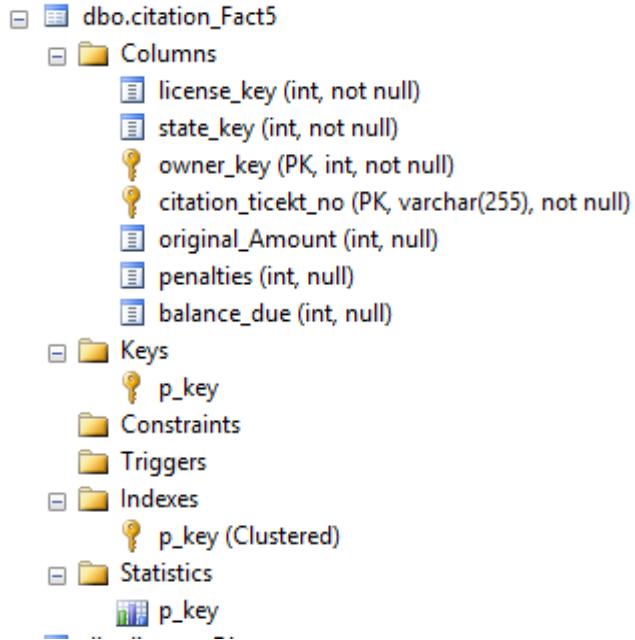
High Level Source to Target Map:



IMPLEMENTATION:

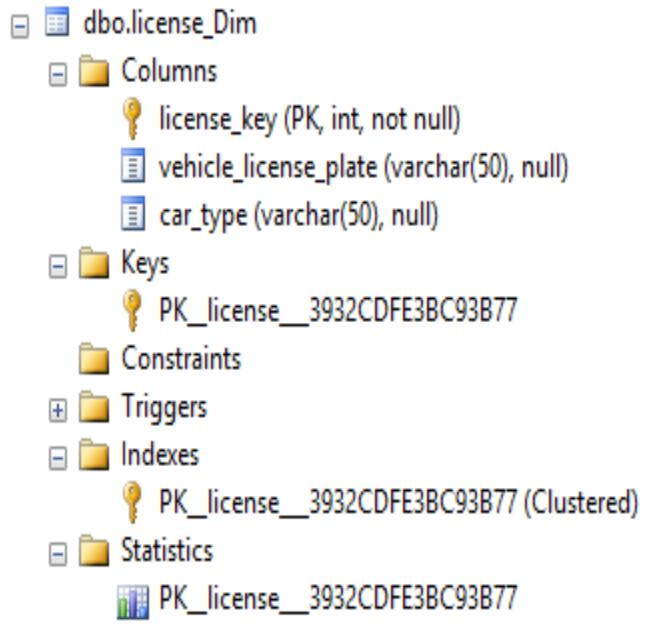
4. ROLAP SCHEMA IN SQL SERVER:

Fact Table: **Citation_Fact**

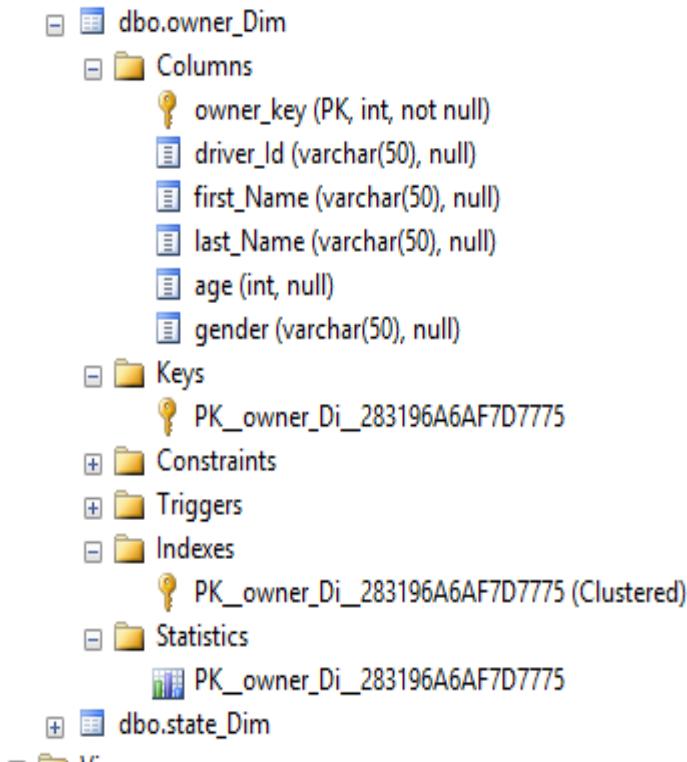


Dimension Tables:

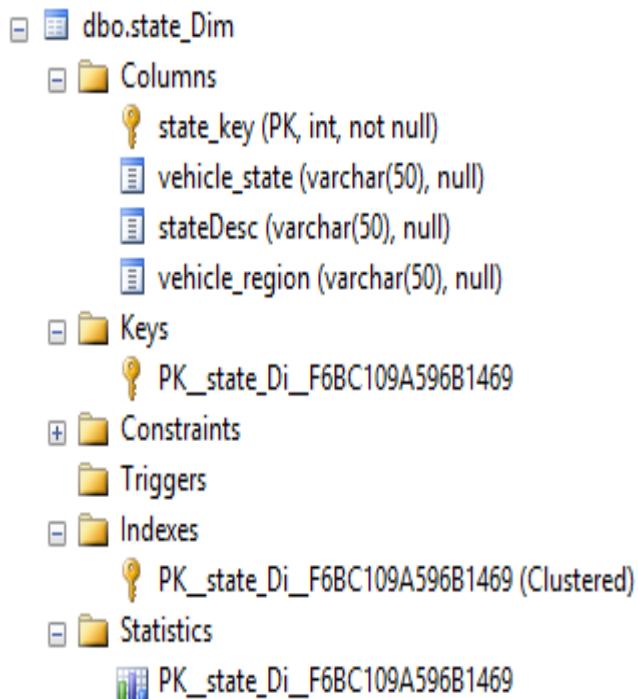
- **DimLicense:**



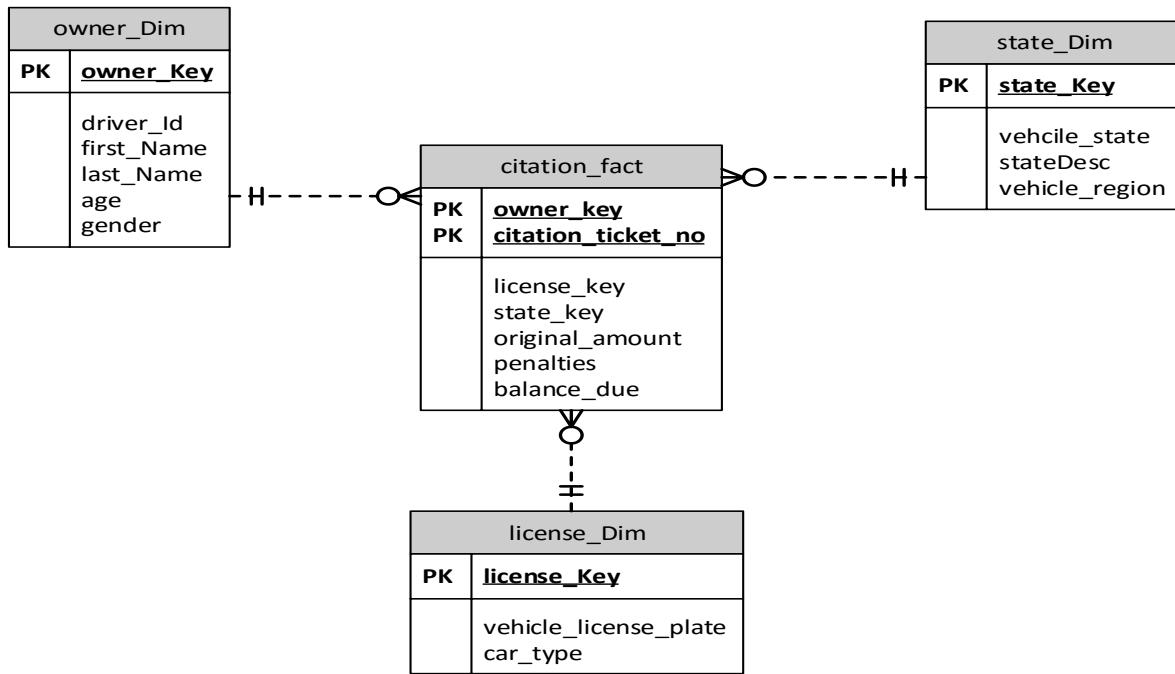
- DimOwner:



- DimState:

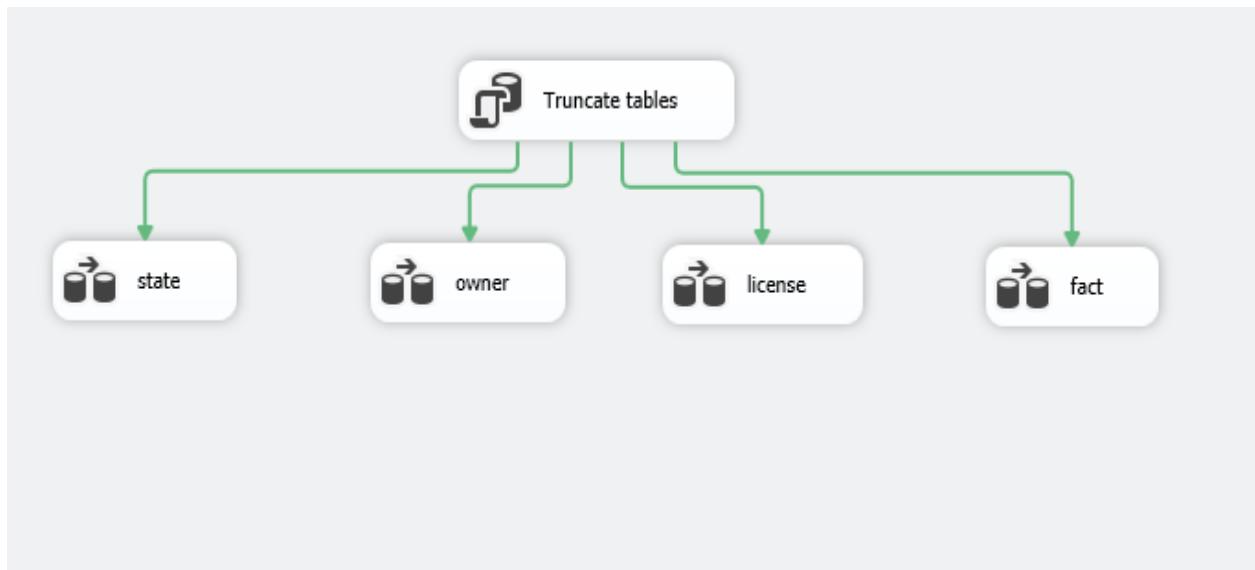


STAR SCHEMA DIAGRAMS

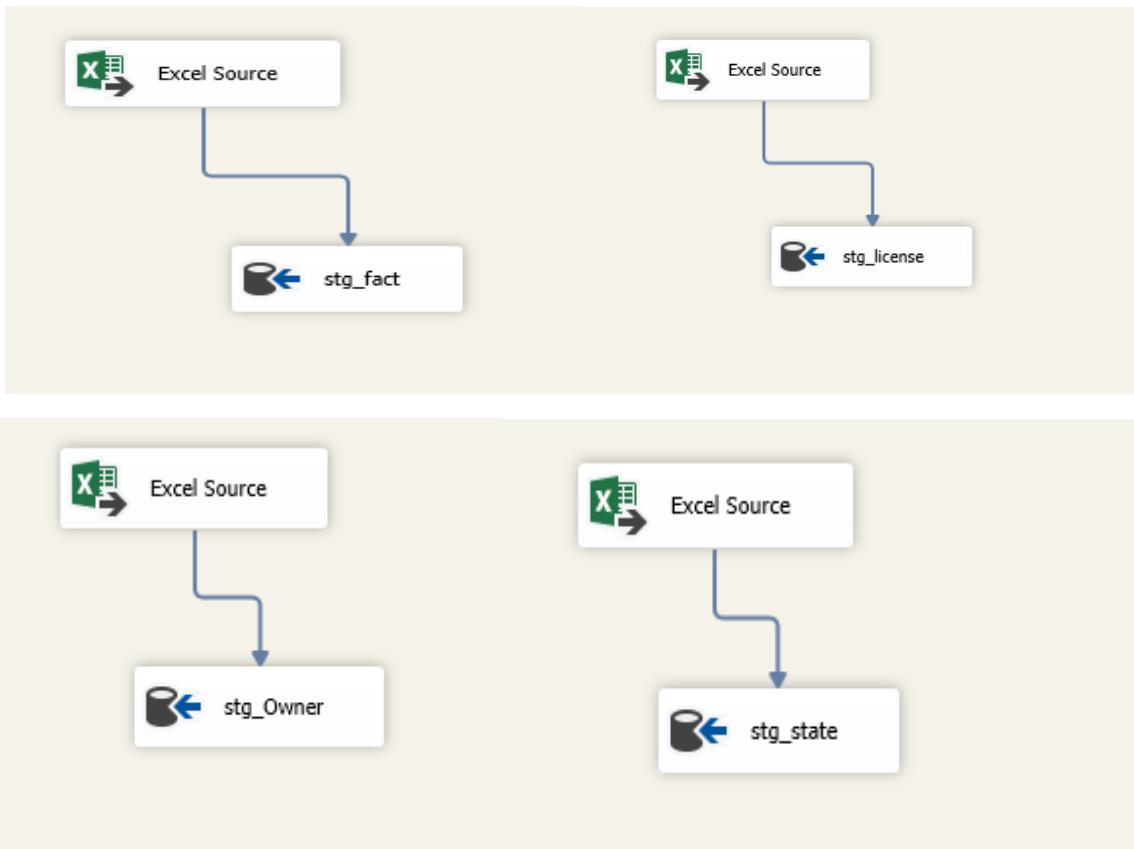


5. SSIS ETL CODE/PACKAGES

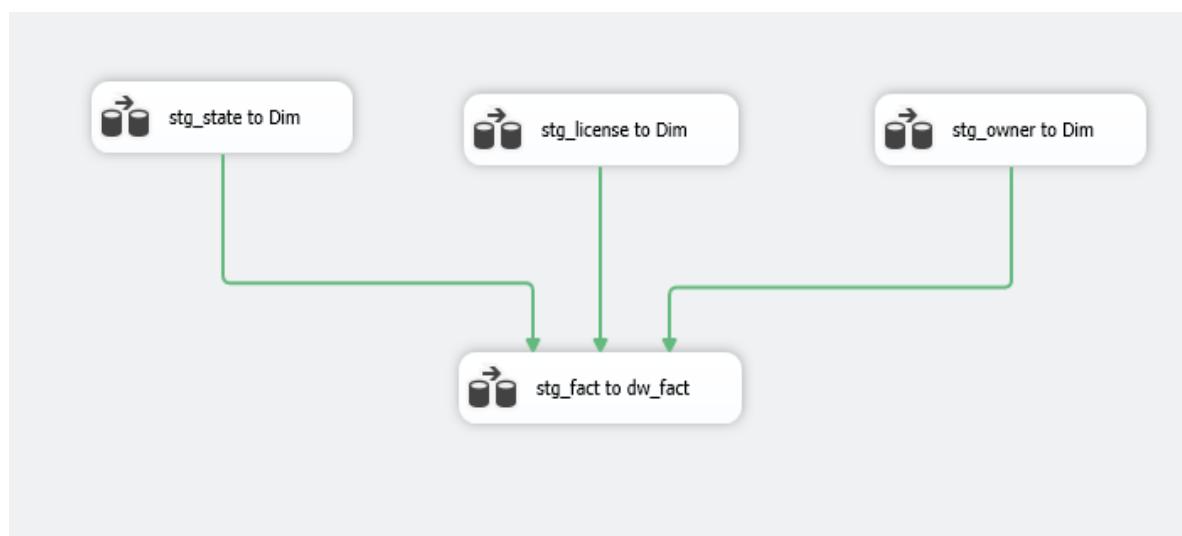
Staging:



In the above diagram, we are truncating all the tables that is used in the project. Here, the truncate table will execute a SQL command and delete all existing rows that are present in the stage tables.

Loading:

It can be seen from above that the various columns that are present in the Source table (in this case, Excel Worksheet) is being considered and appropriate columns are extracted and transferred to the corresponding stage tables.

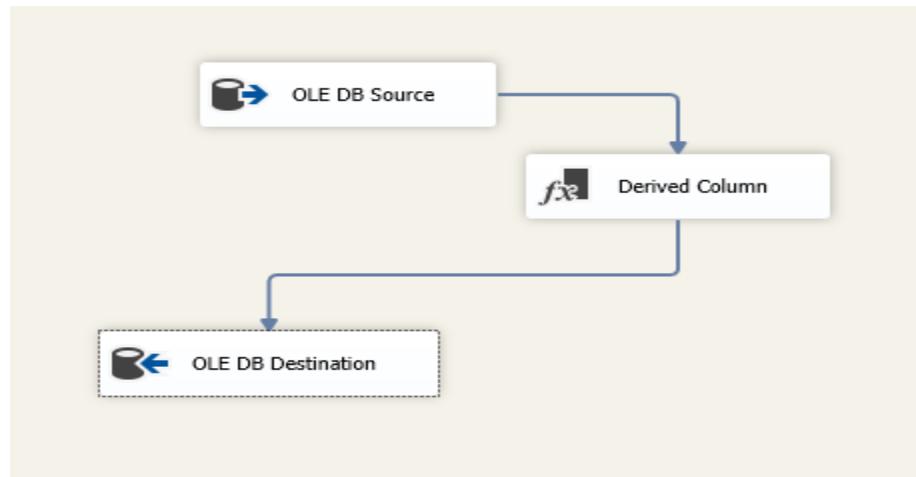


The above figure shows that the various stage tables are being transferred to the Data Warehouse tables. Both the dimension tables and the fact table are transferred in this case.

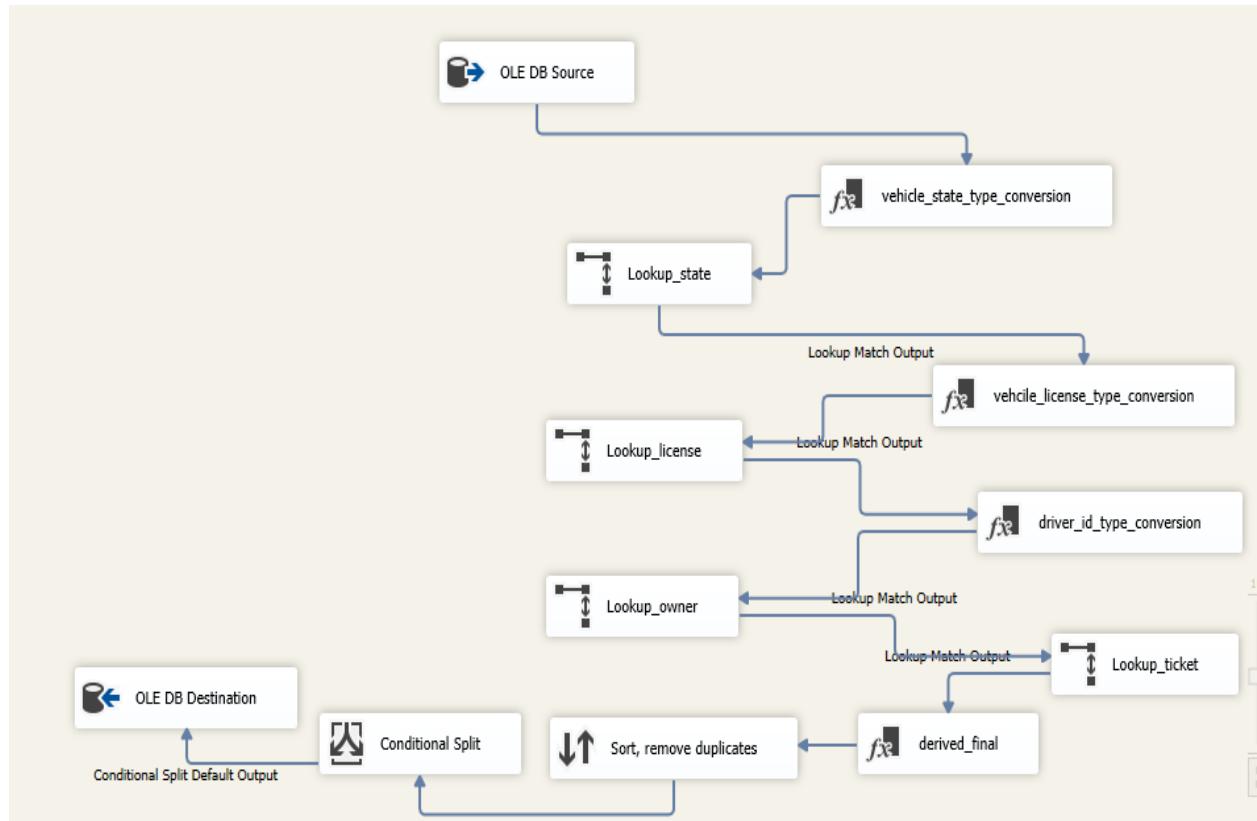
SUBSEQUENT LOADS:

Stage to Dimension Data flow:

In the below figure, the process that occurs when the stage tables are being transferred to the DW tables. Here, the columns that contains Unicode values are being converted into the compatible datatypes, before being transferred to the Data Warehouse.



Stage Fact to Warehouse:

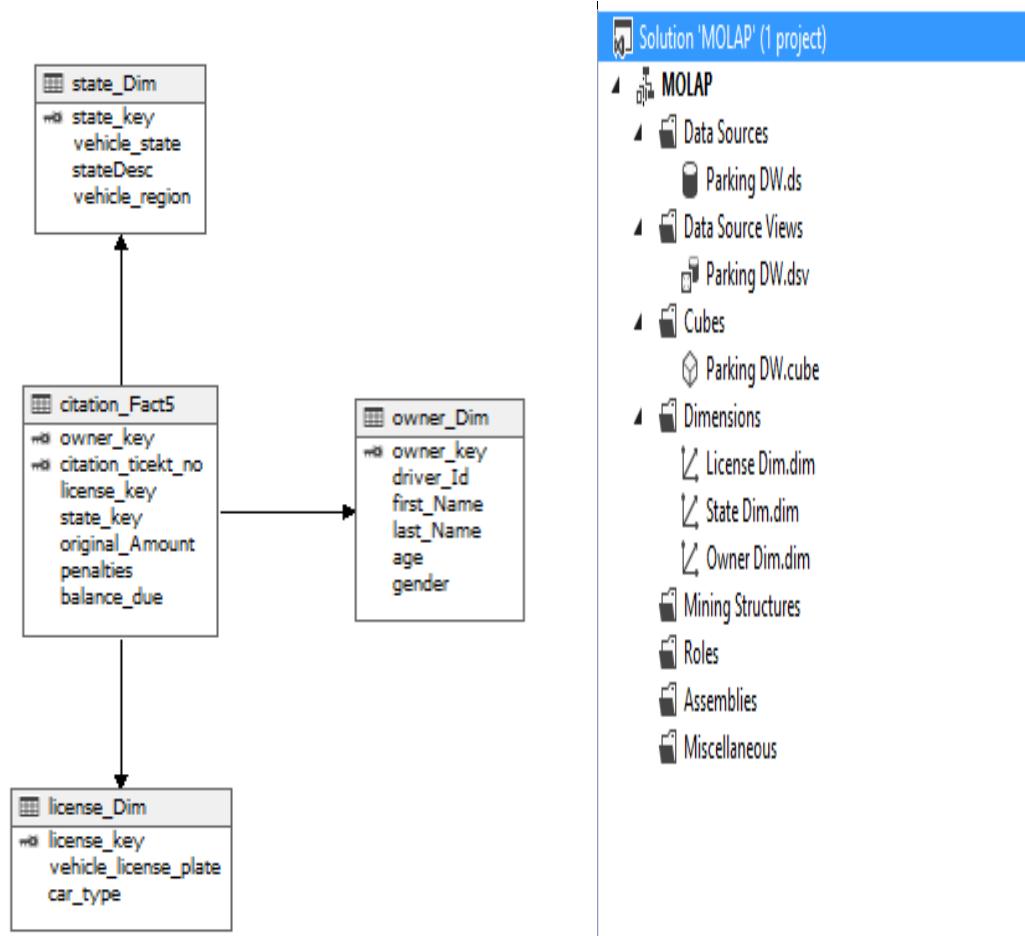


In the above diagram, the following process takes place:

- *Dim_type_conversion and Lookup:* This process extracts the tables in the Stage environment and converts into the compatible datatype in the DW environment. It also looks-up the appropriate keys in the DW environment (State_Key, License_Key, Owner_Key) and matches it with the corresponding ID. Then, only the matched records are extracted and transferred for further process.
- *Sort, Remove Duplicates:* Here, the matched records along with the Citation Ticket No is considered and is sorted based on the latter. Also, here, the duplicate records with the same Citation_No are removed and processed.
- *Conditional Split:* In this step, the NOT NULL values are only extracted from the previous step and is used for the further DW process.

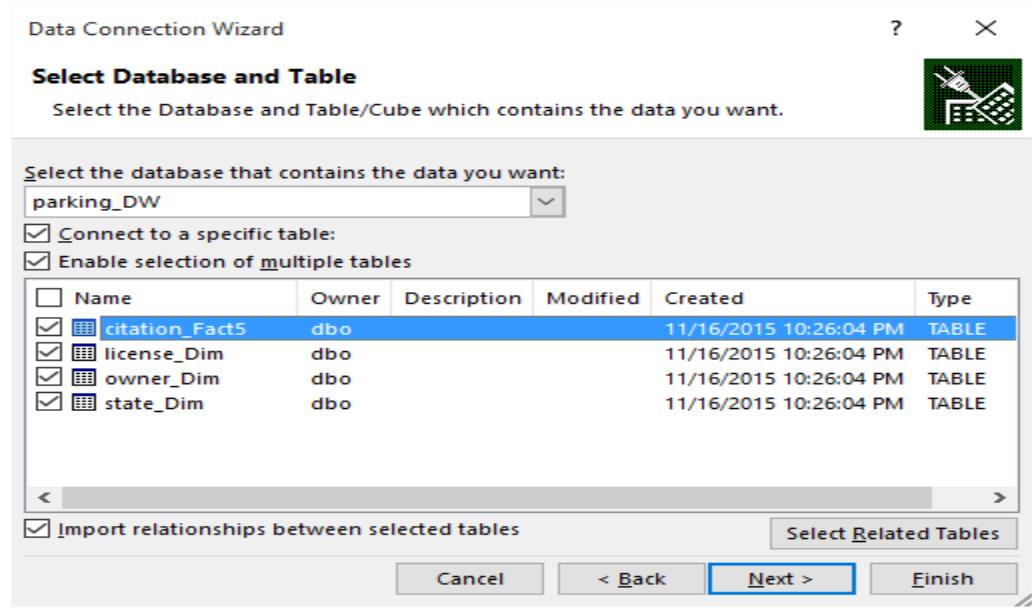
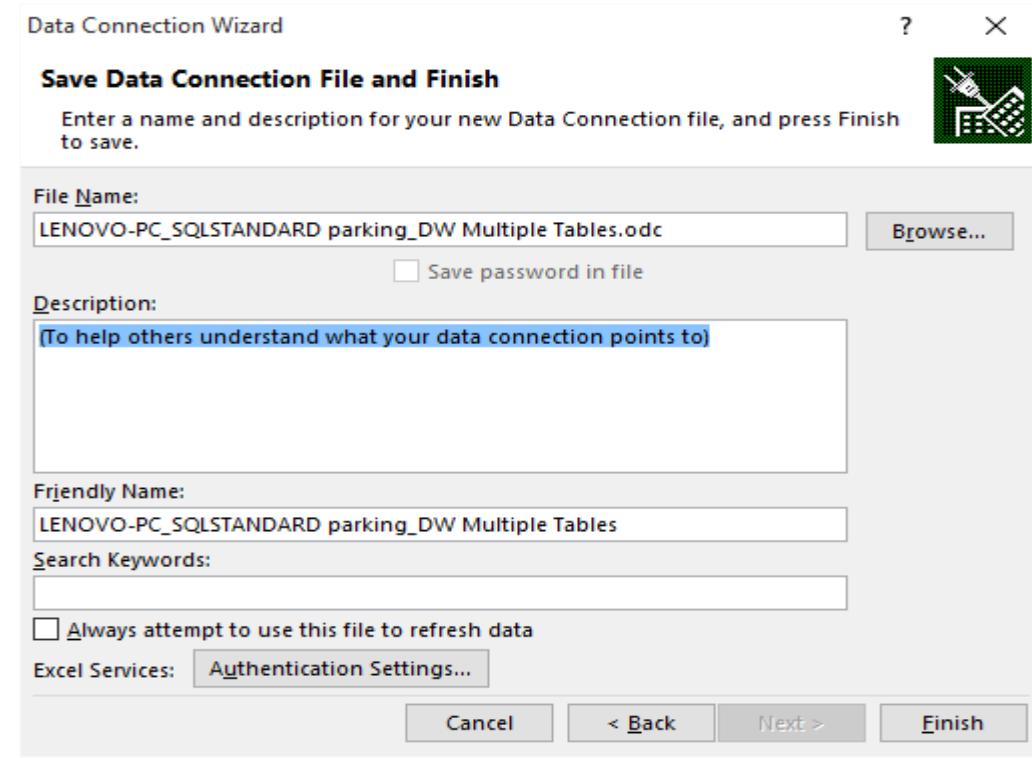
6. Analysis Services MOLAP Database:

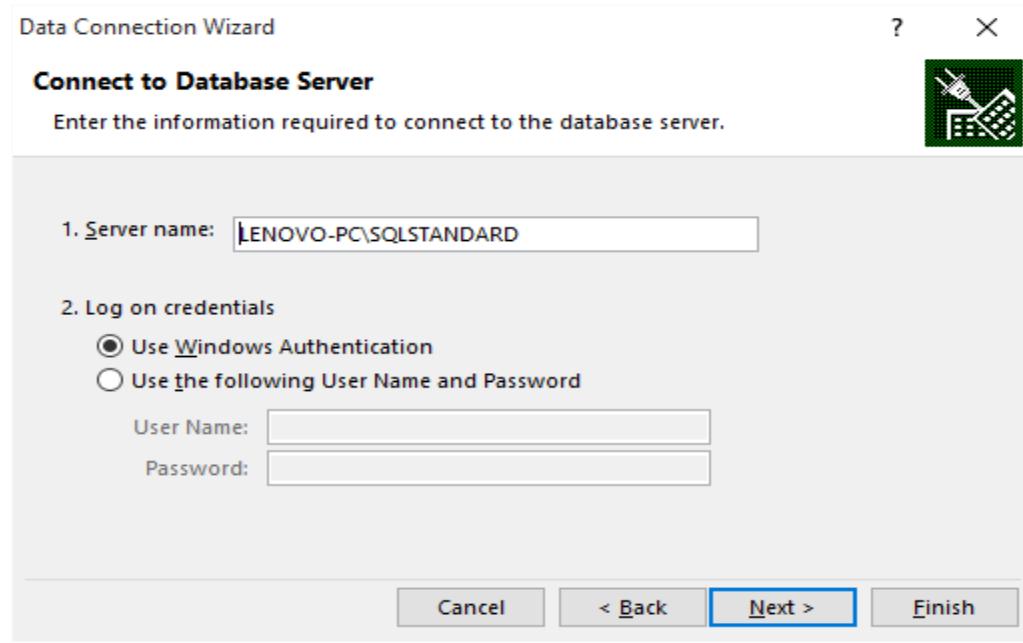
Cubes to match Star Schemas/Business Processes



ANALYSIS AND REPORTING

7. Front-End Linkage





Status	Table ▾	Related Lookup Table	New...
Active	citation_Fact5 (license_key)	license_Dim (license_key)	Edit...
Active	citation_Fact5 (owner_key)	owner_Dim (owner_key)	Activate
Active	citation_Fact5 (state_key)	state_Dim (state_key)	Deactivate

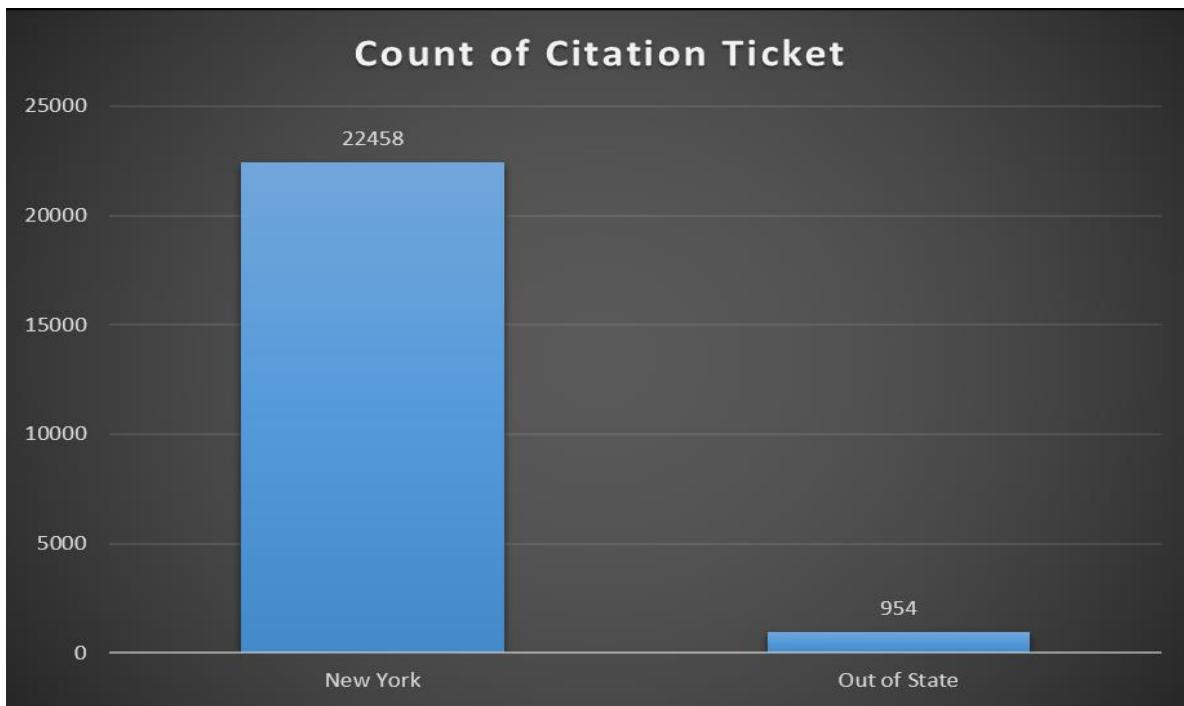
On the right side of the table, there are buttons for 'New...', 'Edit...', 'Activate', 'Deactivate', and 'Delete'. At the bottom right is a 'Close' button.

8. Observation and Analysis:

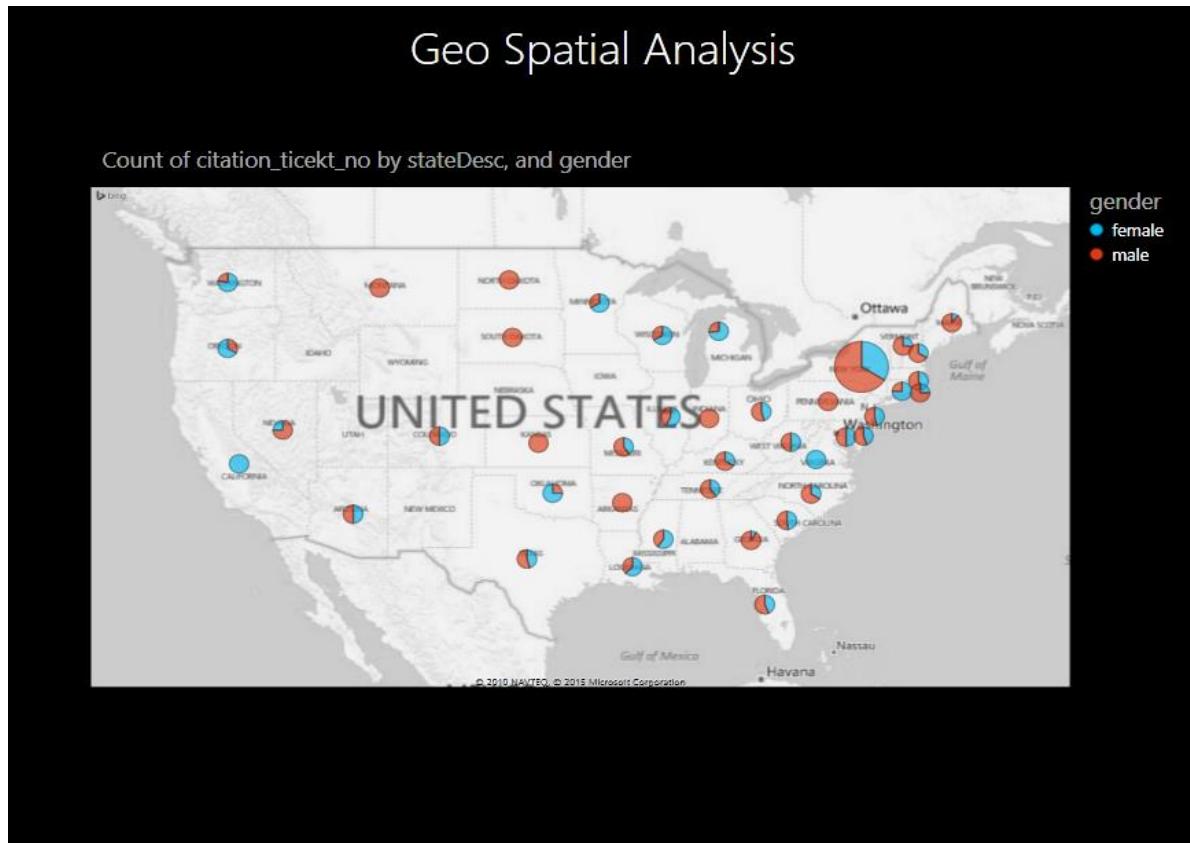
The business objective of the project is to discern patterns from the outstanding parking violations based our analysis of the dataset provided by the open data team of the city of Albany, NY. Interactive dashboards have been constructed after in-depth exploratory analysis of the data set. The team was interested to look out for trends in demographic of the vehicle driver information such as age groups, geographies or gender information played an influencing role in the number of citations obtained as well as total dollar amount value that is outstanding.

As the city of Albany is located in the state of New York as well as it's the state capital, the number of citations received by the in-state vehicles are obviously high. Hence we have segregated the data based on in-state (from the State of New York) and out of state (all other states) status of the vehicles and based our analysis on this primary factor.

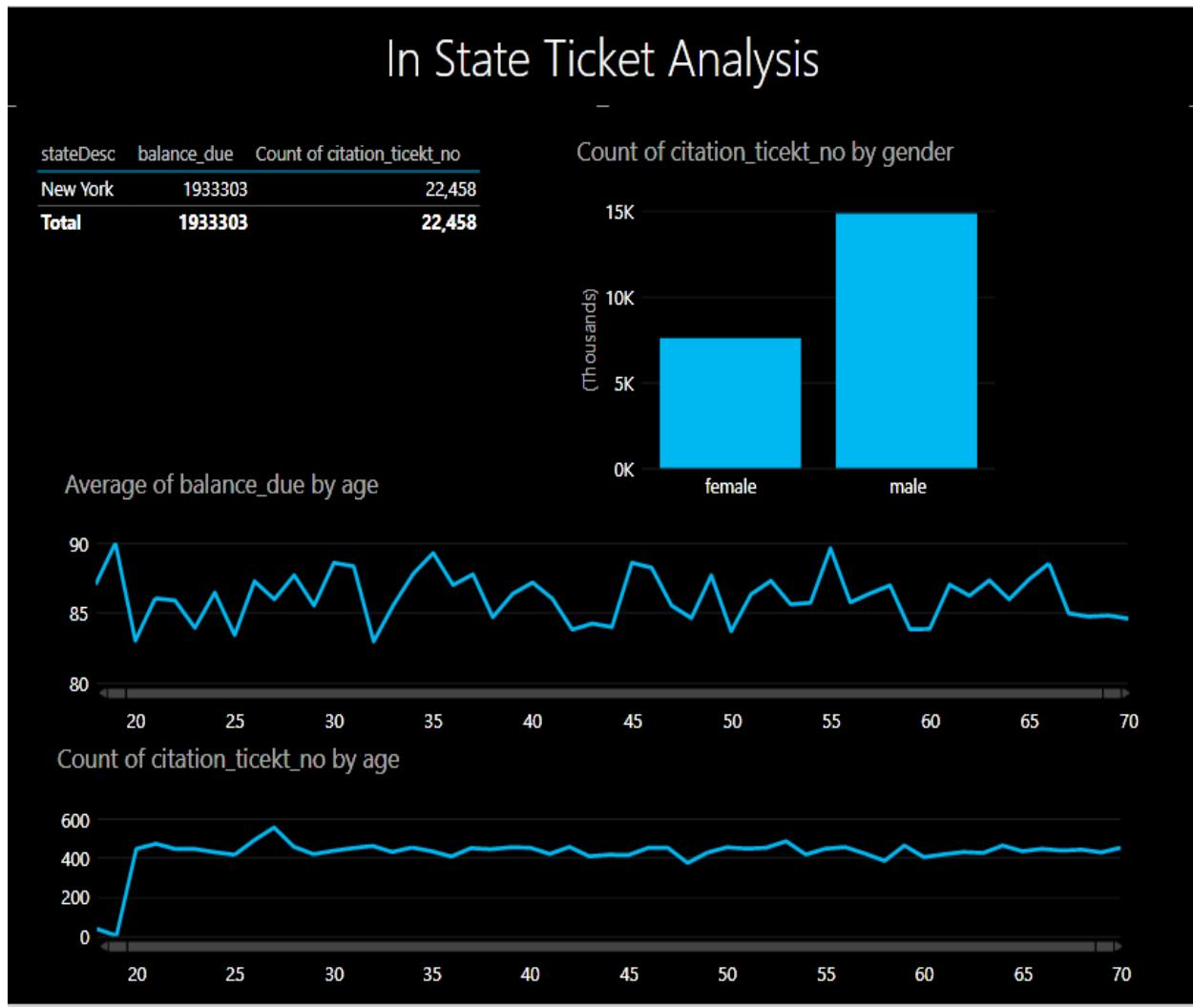
We shall discuss these different analyses and the inferences that can be drawn from the visualized data as well as the actionable insights that might be taken away to identify the root cause or patterns.



A comparison has been made purely based on the number of citations that has been doled out to gain perspective. Despite it being in-state, the number of violation caused by the state of New York are extremely high close to 96%.

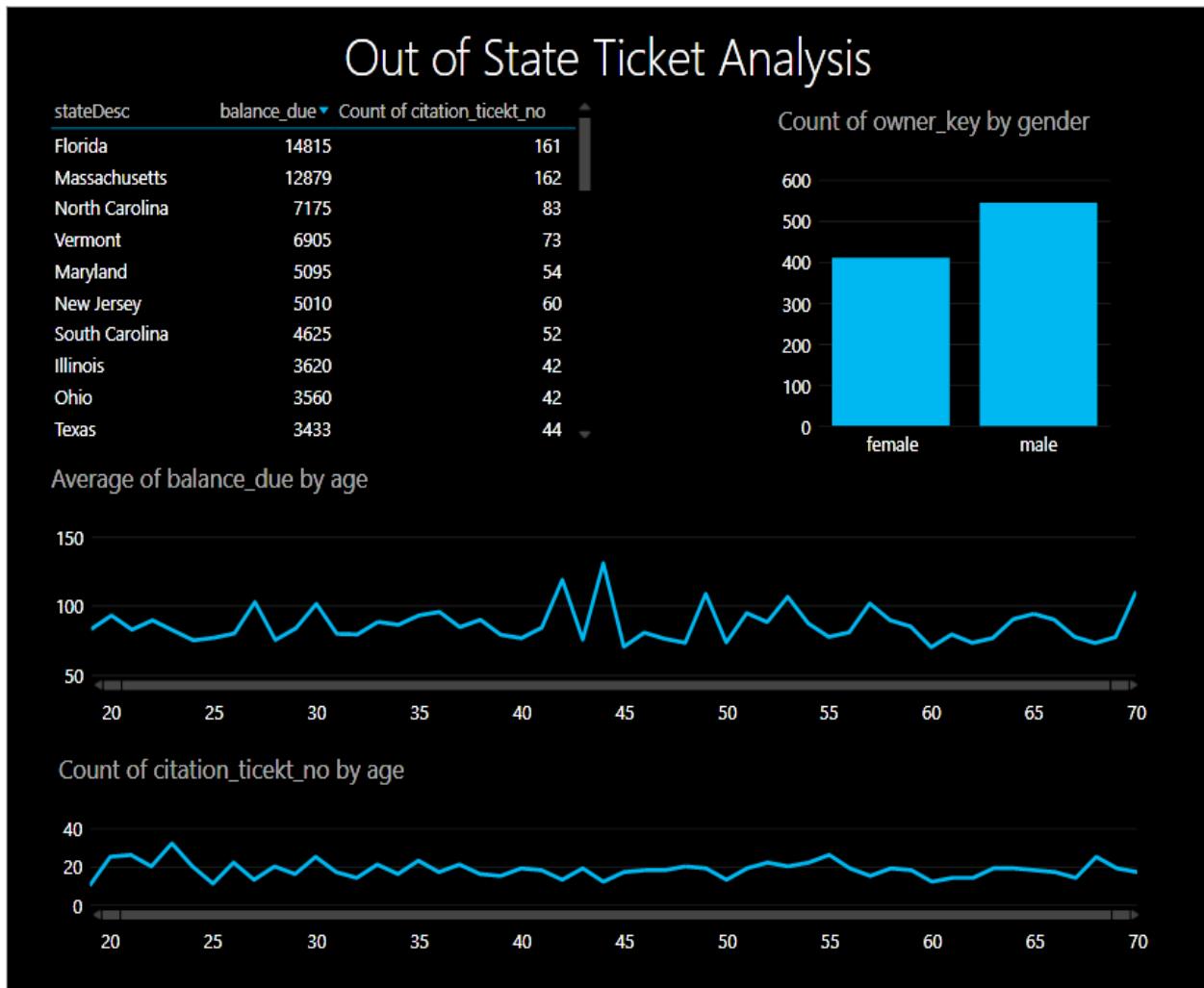


The geographical information retrieved from the license plate number attribute of the data set was superimposed on a map of North America to visualize current state of issue to the officials from the city of Albany along with gender distribution of each state. This graphic may prove to be useful to form a strategy to retrieve the outstanding balance amount from states with the highest outstanding penalties as well as devise methods to mitigate such occurrences in the future by forming ties with the respective state DMVs.

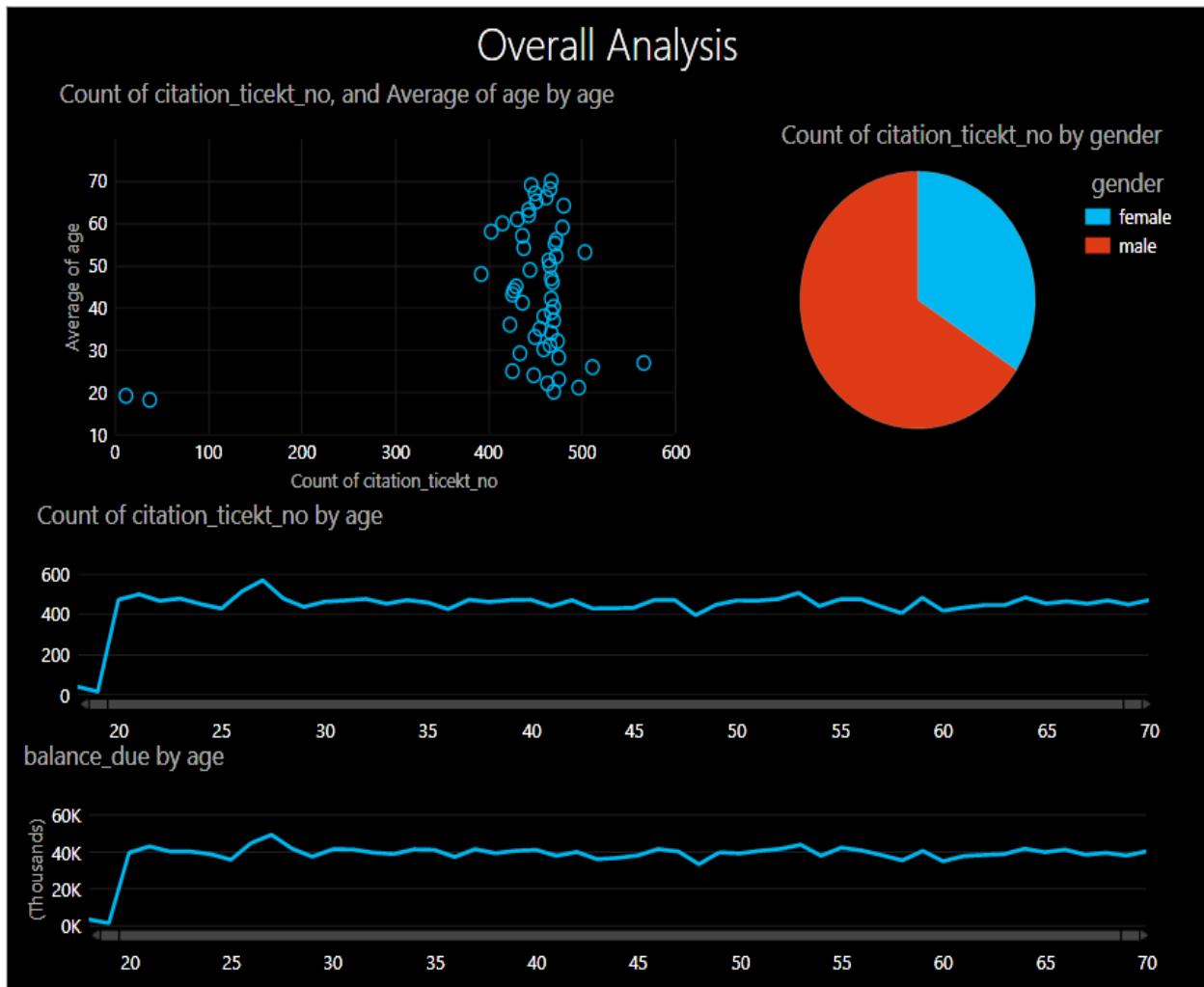


From the available data we observe that the New York State license plates amount to 22,458 citations with result in fines to the tune of 19 million dollars. The distribution of age when compared with the overdue total dollar amount balance was observed to be higher in drivers under the age of 20 and spikes higher around 35 and 55 year olds.

It can be observed that female drivers are performing better in-terms of having relatively lower outstanding citations as compared to their male counterparts at least in the city of Albany.



From our analysis, it is evident that People from Massachusetts have received the highest number of tickets, followed by Florida which is very far from Albany and there could be various changes in terms of regulations and hence more number of tickets issued by the New York State. The amount of balance due is proportional to the number of citation tickets issued.



The overall analysis shows that the number of citation tickets received by male is approximately 1.5 times the tickets obtained by female. It is evident from our timeline graphs that the average age group of the population receiving the tickets lie between 20 and 40 and the highest number of tickets have been issued to people in mid-20s.

The number of citations received is seen to dip sharply around the age group 20 and rises in the mid-30s. Hence while sending notifications to remind these defaulters, officials can form a stratified approach to use available resources effectively and collect the outstanding amount of fines.

ANNEXURES:

Source Data Set:



parking_tickets_fin
al.xlsx

**As the available dataset lacked attributes that could showcase different facets of the business problem and to make the project challenging in terms of architecture and design; the team decided to simulate certain attributes (Gender, Car Type, Region, Age) of the data that could be used for the desired analysis using R Programming and Packages.

Website Link: <https://data.albanyny.gov/City-Finances/Outstanding-Parking-Tickets-11-17-15/9ui3-cap9>

Data Warehouse Script:



dw_script.sql

SSIS Integration Packages:



Package-loading.dt
sx



parking-extraction.
dtsx

Dashboard:



Group3_Project_Da
shboard.xlsx

** Dashboard requires PowerView Excel Add-in to function properly