SJSU Parking Database Management Report



Department of Applied Data Science

San Jose State University

FALL 2019- BUS 243

Group2 Members:

Jeehee (Ginny) Choi

Prathusha Koouri

Aparna Menon

Sukriti Mishra

**Table of Contents**

[***Project Overview***](#_heading=h.30j0zll) 3

**Introduction** 3

**Objectives** 3

**Analytical Requirements** 4

**Data Overview** 4

[***Data Model Description and ER diagram***](#_heading=h.3dy6vkm) 6

[***Query and Report Needs of Organization***](#_heading=h.1t3h5sf) 9

[***Conclusion***](#_heading=h.4d34og8) 19

[***References***](#_heading=h.2s8eyo1) 20

# Project Overview

## Introduction

We chose SJSU (San Jose State University) Parking Database Management System’ as our group project topic because one team member has frequently experienced parking difficulties on-campus and became eager to analyze the root causes as a data analytics major. She purchased an on-campus semester parking permit, but she had to pay additional private parking garage fees or wait over an hour in parking garages because on-campus student parking lots reached full capacity.

To understand the current status and challenges of the SJSU parking system, our team had face-to-face and phone interviews with SJSU Parking Services managers and researched SJSU parking policies, campus master plan for parking, and news articles about future parking plans. Through these interviews and researches, we found possible causes of parking lot challenges.

While SJSU has a total of 36,000 students and 4,300 employees (San Jose State University, "About SJSU", 2019), “the Main Campus has .38 parking spaces per full-time equivalent (FTE) student… below the nationally recognized average of .50 per FTE...” (San Jose State University, "Master Plan 2001", 2001, pp. 2-13). This indicates that the SJSU population has been increased while SJSU parking spaces have long been a concern.

We analyzed possible root causes of parking lot challenges by setting objectives and analytical requirements of this project, then determined which data to collect for further analysis.

## Objectives

We focused on five objectives in this project.

1. Find SJSU parking challenges:

* Lack of parking spaces for students
* Preference for certain parking lots over others
* Low turnover of cars in lots.

1. Research SJSU parking infrastructure, permit types, and operations.
2. Identify entities in the parking database and establish relations among the entities.
3. Diagnose root causes of parking challenges using SQL queries and visualize the data using Tableau for better understanding.
4. Plan future scopes for further real-time parking data analysis.

## Analytical Requirements

Below are the seven analytical requirements in this project.

1. Acquire publicly available SJSU parking data as much as possible.
2. Generate private data using the Mockaroo website.
3. Create an ER diagram and generate a database.
4. Normalize and cleanse data and set primary/ foreign keys.
5. Determine SQL database table relationships (one-to-one, one-to-many, many-to-many, etc.)
6. Create meaningful SQL queries corresponding to SJSU parking challenges.

* Identify SJSU parking customers.
* Classify permit holders in each category.
* Explore the most preferred parking lots by different permit holders.
* Detect overstaying cars by permit holders.
* Find the most frequently-issued citations.

1. Draw explicit insights from SQL query results.

## Data Overview

We acquired publicly available SJSU parking data from the SJSU Parking Services website, generated dummy data for private data using Mockaroo, and created a prototype of the SJSU parking database management system.

The data we collected or generated are:

1. Permit price by type:

* Student permit price by type ("Student Commuter Virtual Permits", 2019)
* Employee permit price by type ("Employee Permits", 2019)
* Housing permit price by type ("Housing Virtual Permits", 2019)
* Park & Ride lot permit price by type ("Housing Virtual Permits", 2019)
* Visitor/ Guest parking permit price by type ("Visitor/ Guest Parking", 2019)
* Motorcycle Permit price by type ("Motorcycle Permits", 2013)

1. Citation penalty: ("Citation Fines", 2013)
2. Parking Lot:

* South/West/North parking garages ("Parking Garages", 2019)
* On-campus parking lots ("Parking Lots", 2019)
* Additional parking lots ("Additional Parking", 2019)

1. Dummy data generated using Mockaroo:

* Customers data (Koouri, "Customer Citation Data Schema", 2019)
* Occupancy data (Choi, "Occupancy Data Schema", 2019)
* Customer citation data (Choi, "Customer Citation Data Schema", 2019)

# Data Model Description and ER diagram

Depending on the needs of the facility, a parking system database could be used to house multiple information. Some of this information will be static information like parking lot information (name, location, geo-coordinates, description), which would have to be entered on-time manually into the database. There will also be a lot of dynamic information such as real-time data from sensors, monitoring the occupancy of the parking lot.

In this project, we designed the SJSU parking database keeping in mind the functionalities of the University parking facility. Our database enables a parking system official to track customers, their permit type, the occupancy rates of the parking facility, and percentage turnover over any given time. The database also enables tracking of citations that occur in the parking facilities.

For our model, we used MySQL to develop an ER diagram and tables for the database. We chose six different entities for the database, as illustrated in Figure 1. The ER diagram and the description of the entities and their attributes are provided below.

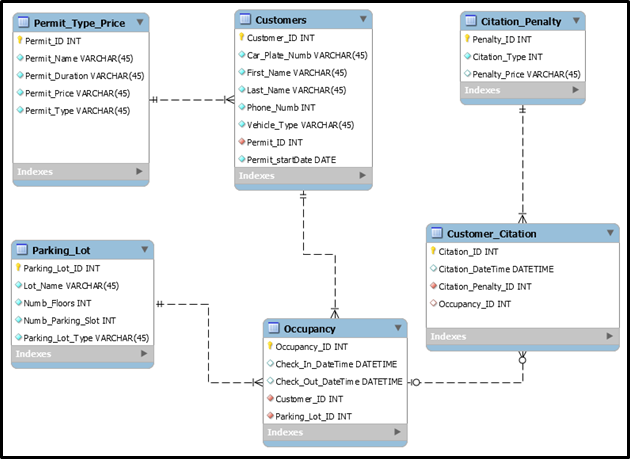


Figure 1 Entity-Relationship Diagram for SJSU Parking Database

Customer Table: This table consists of personal data of customers such as customer name, phone number, vehicle type, permit id and start date of permit, etc., registered in the University parking website to purchase the parking permits. For every registered customer, a unique customer id is generated, which serves as the primary key for the entity. The type of permit purchased by the customer is identified by the permit id, which acts as a foreign key, connecting the Permit\_Type\_Price entity to the customer table. The following are the attributes in the customer table:

Customer\_ID, Car\_Plate\_Numb, First\_Name, Last\_Name, Phone\_Numb, Vehicle\_Type, Pemit\_ID and Permit-startDate

Permit\_Type\_Price table: This table is a static table, containing the various attributes of permits. Each permit has a unique id, acting as the primary key. We have included 23 different permits under student, employee, and visitor categories. Various attributes of the permit\_type\_price table are:

Permit\_ID, Permit\_Name, Permit\_Duration, Permit\_Price, Permit\_Type.

Parking\_Lot table: This table consists of the attributes of the parking lot. Like permit\_type\_price table, this table is also a static table. Each parking lot is identified by a unique id, which serves as the primary key of the entity. The attributes of the Parking\_Lot table are as follows:

Parking\_Lot\_ID, Lot Name, Numb\_Floors, Numb\_Parking\_Slot, Parking\_Lot\_Type

Occupancy table: This table gets updated every time a customer checks in or checks out of the parking lot. In real-world situations, a sensor monitors the entry and exit of the vehicle from the parking lot, updating the database every time. Each occupancy in a parking lot by a customer generates a unique occupancy id, which acts as the primary key for the entity. This table has two foreign keys: Customer\_ID and Parking\_Lot\_ID, connecting the table to both Customers table as well as Parking\_Lot table. Following are the attributes of the occupancy table:

Occupancy\_ID, Check\_In\_DateTime, Check\_Out\_DateTime, Customer\_ID, Parking\_Lot\_ID.

Customer\_Citation table: This table is updated every time a citation occurs. Whenever a citation occurs, a citation id is generated, which acts as the primary key for the entity. Each citation relates to the Citation\_Penalty table through the foreign key Penalty\_ID. In case the violation is caused by a registered customer, occupancy id connects that record to the occupancy table. The Customer\_Citation table has the following attributes:

Citation\_ID, Citation\_DateTime, Citation\_Penalty\_ID, Occupancy\_ID

Citation\_Penalty table: The table entails 25 types of citations with citation amounts in the University. Each citation has a penalty id, which serves as the primary key for the table. Following are the attributes for this table:

Penalty\_ID, Citation\_Type, Penalty\_Price

# Query and Report Needs of Organization

As per the stated problem of limited parking spaces at the San Jose State University. We analyzed the problem and found that there is a lot of sensitive information in the original parking data, which is not available in full for analysis. Then, we decided to create/generate mock data in Mockaroo website for the tables specified in the ER Diagram given above.

After creating database and tables by forward-engineering in MYSQL, we generated mock data in the form of SQL insert statements for all the tables accordingly. We directly executed statements to fill tables in the database with the generated Mock Data.

To fulfill the analytical requirements of the organization mentioned in the above section, we decided to query the database to gain meaningful information/insights. The following are the sample queries we generated with the output.

* Our first query is to find out the number of permits issued each year for all the entities. It is based on customers and permit\_type\_price tables. We joined these two tables and extracted information. Below is the query along with the output:

Query 1: How many permits are issued for students, visitors, and employees each?

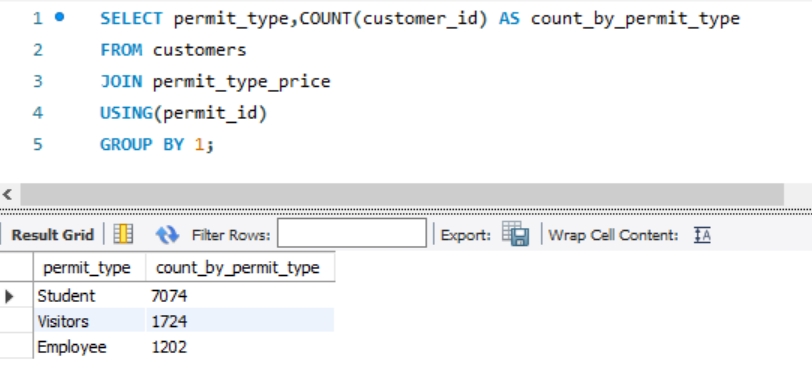
SELECT permit\_type,COUNT(customer\_id) AS count\_by\_permit\_type

FROM customers

JOIN permit\_type\_price

USING(permit\_id)

GROUP BY 1;



From the above query output, we see that three types of permits exist on campus, and the number of permits issued for students is way higher than that of visitors and employees. The permits sold to students each year are about seven times when compared to employees.

* The next query is to analyze the type of vehicles used by different permit holders. This query is intuitional and gives information about the number of vehicles along with their types. We can make sure when designing lots in the future to keep track of the type of vehicles as well.

Query 2: What are the different vehicle types used by permit holders?

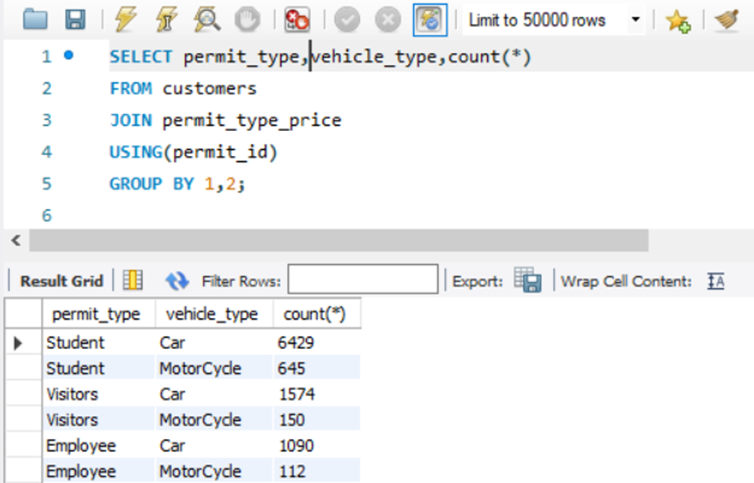
SELECT permit\_type, vehicle\_type, COUNT(\*)

FROM customers

JOIN permit\_type\_price

USING(permit\_id)

GROUP BY 1,2;



From the output of the above query, some people commute with motorcycles, too. In

each permit type, there are about 10% of vehicles as Motorcycle. These vehicles can save

lots of parking space and can be parked in very compact spaces as well. This analysis would help us efficiently using parking lot spaces eventually.

* The third query is to analyze the type of permits issued. By analyzing such details will help us know more about the data.

Query 3: What is the number of permits issued by permit names?

SELECT permit\_type, permit\_name, COUNT(customer\_id) AS count\_by\_permit\_type

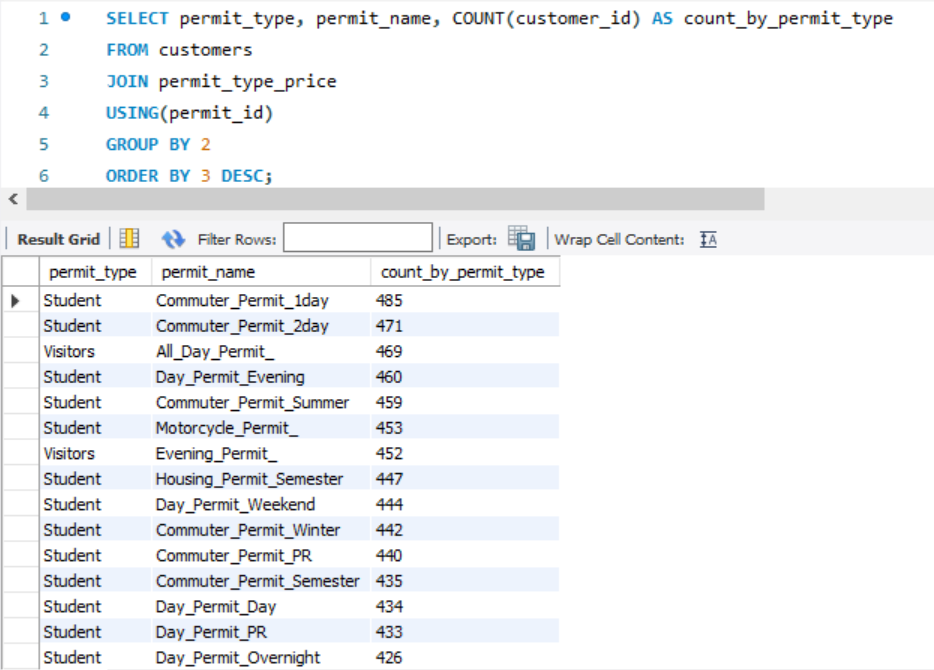
FROM customers

JOIN permit\_type\_price

USING(permit\_id)

GROUP BY 2

ORDER BY 3 DESC;



We see that the names of each permit and the numbers of these permits issued. We also

see that each permit has an approximately equal number of spots allocated. We can also

observe that permits for students have more categories when compared to visitors and

employees.

* We picked a random day and counted the number of incoming vehicles on the day and analyzed their types.

Query 4: How many vehicles are parked on a random day categorized by permit\_type.

SELECT Permit\_Name, Permit\_Type, count(\*) as "No. of vehicles parked on a Random Day"

FROM occupancy JOIN customers USING (customer\_id)

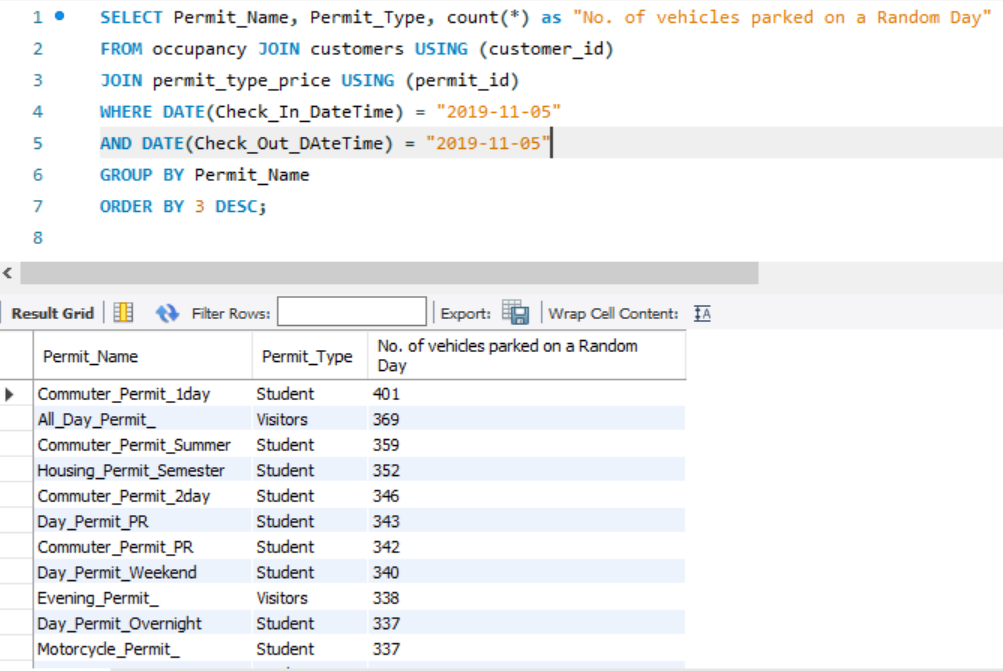
JOIN permit\_type\_price USING (permit\_id)

WHERE DATE(Check\_In\_DateTime) = "2019-11-05"

AND DATE(Check\_Out\_DAteTime) = "2019-11-05"

GROUP BY Permit\_Name

ORDER BY 3 DESC;



From the above query output, we analyzed that the Permit “Commuter\_Permit\_1day” for

the student is the most used permit on that particular day, probably aligned with the class

schedule. “All\_Day\_Permit” holders of visitor category are the second-highest users of

the parking spot, which is normal in November as we see lots of visitors who want to start

their classes in the spring semester plan to visit the university. Besides, we see the

“Commuter\_Permit\_Summer” permit holders, which is out of normal and needs to be

further analyzed.

* The next step is to analyze the distribution of occupied spots in different garages. This is to get the intuition of which garages are most used and which garages have the most incoming and outgoing cars.

Query 5: Which lot is preferred the most to park?

SELECT lot\_name as most\_used\_lot, max(permit\_count) as permit\_count

FROM (select lot\_name as lot\_name, permit\_type as permit\_type,count(permit\_id) as permit\_count

FROM parking\_lot

JOIN occupancy USING (parking\_lot\_id)

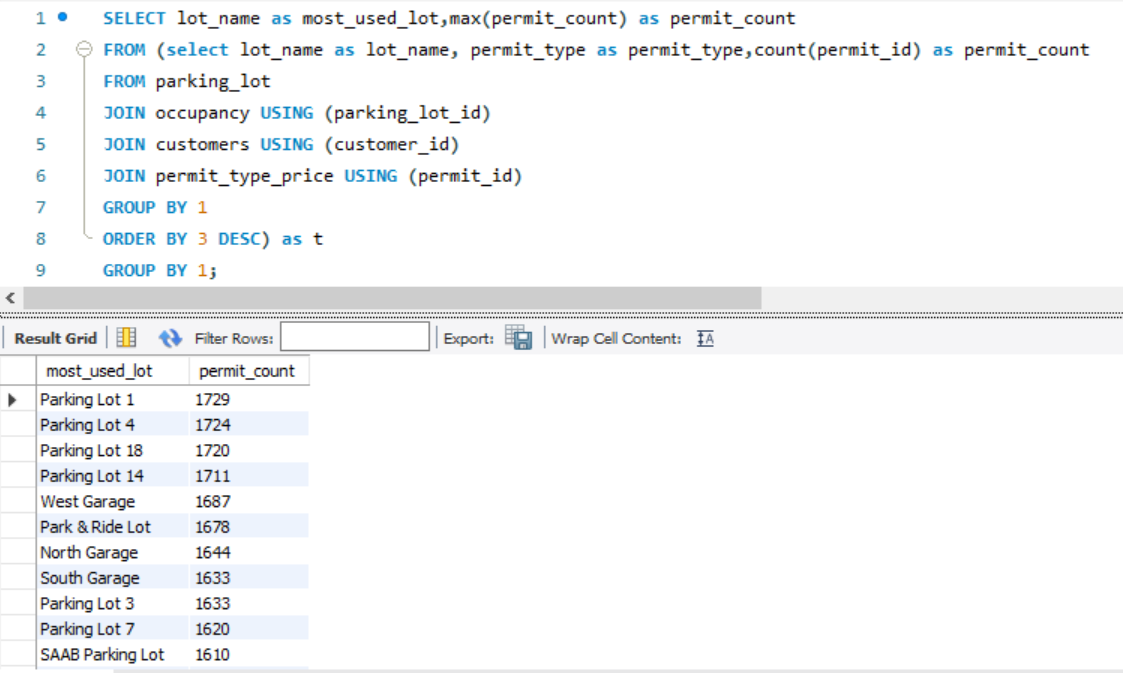
JOIN customers USING (customer\_id)

JOIN permit\_type\_price USING (permit\_id)

GROUP BY 1

ORDER BY 3 DESC) as t

GROUP BY 1;



From the above query output, we see that Parking Lot 1 has the highest turnover of cars

and is the most frequently used by parking customers. Now, let's find out which group

is using Parking Lot 1 a lot.

Query 6: What is the occupancy distribution in different lots by permit type?

SELECT t.lot\_name AS most\_used\_lot,t.permit\_type,max(permit\_count) AS

permit\_count

FROM (select pl.lot\_name AS lot\_name, pp.permit\_type AS

permit\_type,count(permit\_id) AS permit\_count

FROM parking\_lot pl

JOIN occupancy o USING (parking\_lot\_id)

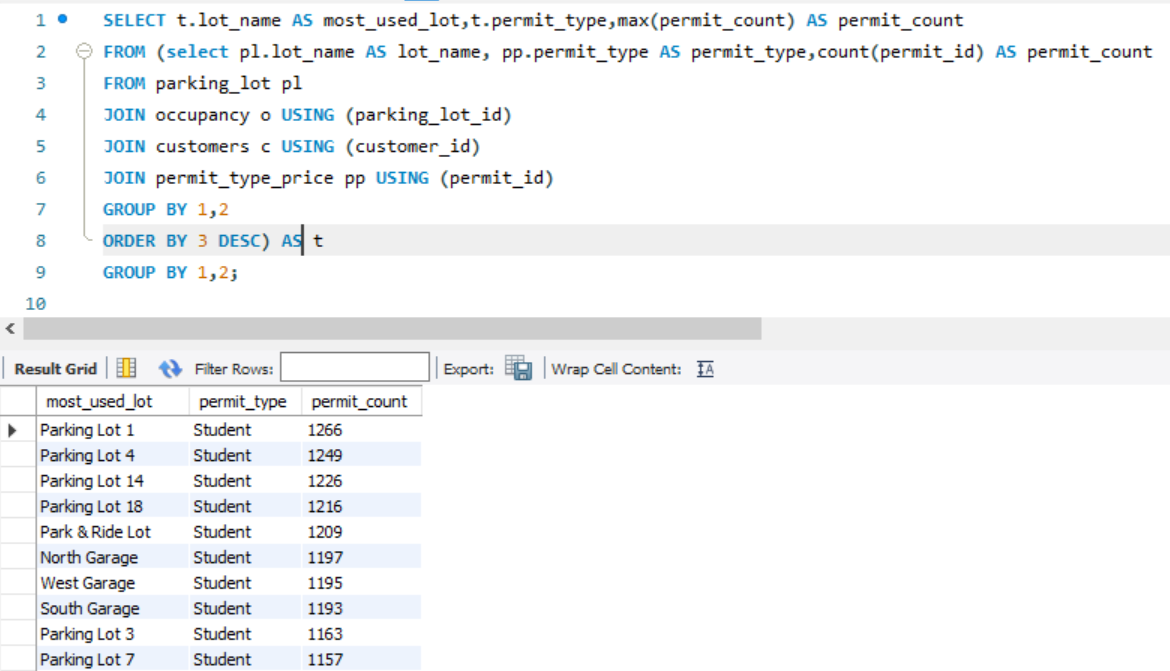
JOIN customers c USING (customer\_id)

JOIN permit\_type\_price pp USING (permit\_id)

GROUP BY 1,2

ORDER BY 3 DESC) AS t

GROUP BY 1,2;



From the partial output above we see that Parking Lot 1 is most frequently used by

students.

* As from the above, we see that a few parking lots are used more frequently and have higher turnover of vehicles. So we further analyzed the vehicles parked for a long time. If those permit holders have other than the housing permit or one-day overnight perimeter permit, it is a parking violation.

Query 6: How many vehicles by permit type that are parked for more than three days in different lots?

SELECT ptp.permit\_type,ptp.permit\_name,pl.lot\_name,COUNT(pl.lot\_name) AS "# of vehicles parked over 3 days"

FROM occupancy o

JOIN parking\_lot pl

ON o.parking\_lot\_id=pl.parking\_lot\_id

JOIN customers c

ON o.customer\_id=c.customer\_id

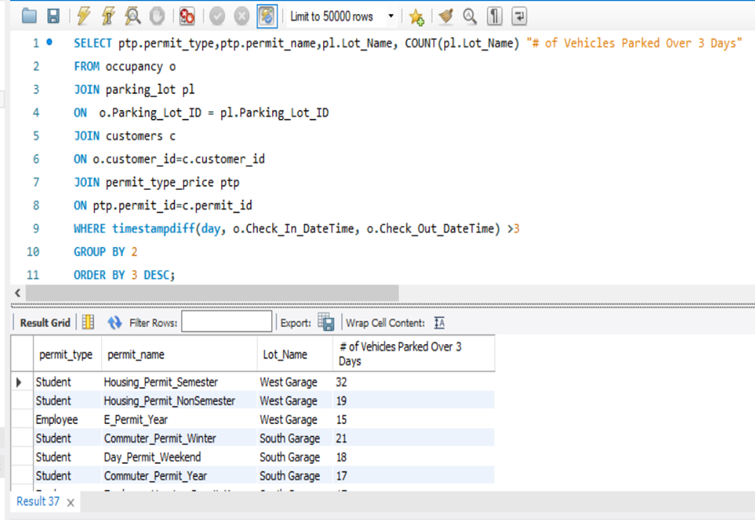
JOIN permit\_type\_price ptp

ON ptp.permit\_id=c.permit\_id

WHERE timestampdiff(day,o.check\_in\_datetime,o.check\_out\_datetime)>3

GROUP BY 2

ORDER BY 3 DESC;



From the above screenshot, we see that the vehicles which are parked for more than three days

are mostly housing permit students, which is normal. However, we also see that there are

anomalies like Day\_Permit\_Weekend holders, parking their vehicles for more than three

days which would normally prompt citation. Another observation that can be made from

the output(though full output is not seen) is that the vehicles parked overnight only exist in

garages and not in an open parking lot which is also a useful piece of information.

* Finally, for this mini-project, we wanted to analyze the citations and their revenue to the university. We specifically analyze the citation revenue as the contributions from this revenue source are used towards the development of alternative transportation systems ("Parking Information", 2019, p. 8).

Query 7: What is the revenue generated by each citation type?

SELECT cp.citation\_type, COUNT(cc.citation\_penalty\_id) AS "# of citations",

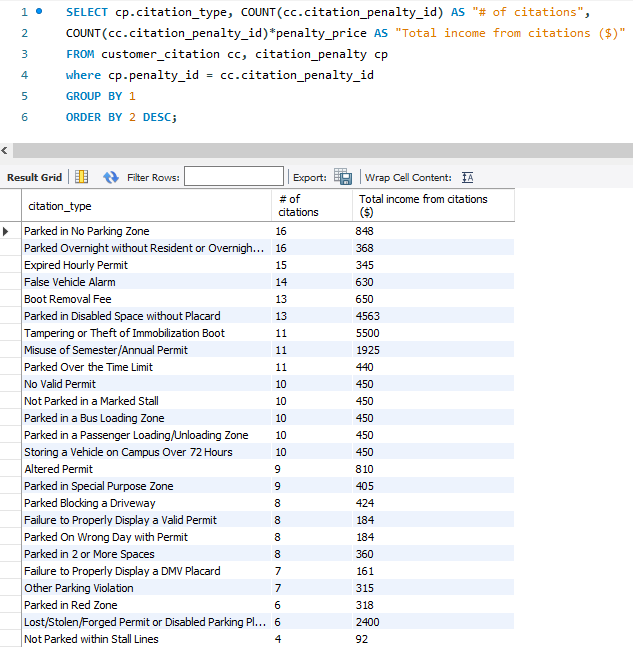
COUNT(cc.citation\_penalty\_id)\*penalty\_price AS "Total income from citations ($)"

FROM customer\_citation cc, citation\_penalty cp

where cp.penalty\_id = cc.citation\_penalty\_id

GROUP BY 1

ORDER BY 2 DESC;



From the above screenshot, we see that “Parked in No Parking Zone” has the most

frequent citation. This could be the case as there are limited parking spaces, and after

getting tired of finding a spot, customers park at ‘no parking zones.’ The next frequently

issued citation is the “overnight parking without a resident permit” which we saw in our

last query that there are recurrent violations of overnight parking.

Below are the similar query results in the Tableau dashboard for further visual analysis. It

helps to have a clear picture of what these query results are.

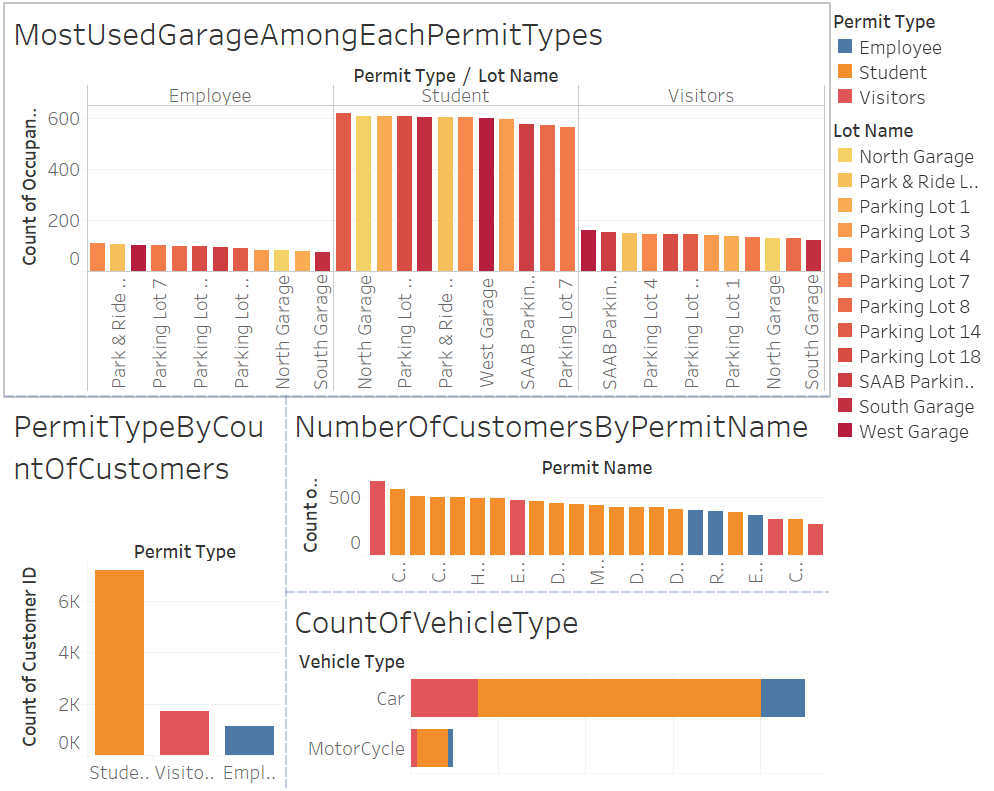


Figure 2 Analyzing Permit Types and Customer Data

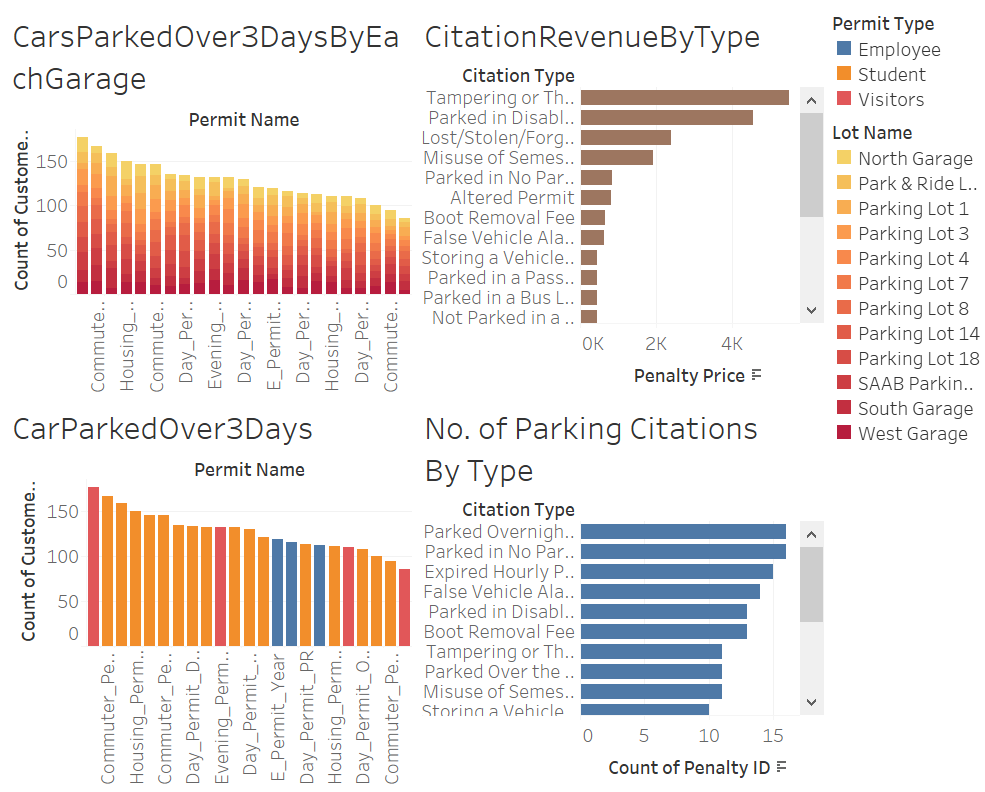


Figure 3 Analyzing Citation Data

# 

# 

# Conclusion

For this project, we were looking for a topic in the area where we face challenges in daily life. The biggest challenge for SJSU students is the lack of parking spaces for students. So, we started our project by selecting the SJSU Parking Management System. It is developed to support the customers and the management of parking with hassle-free experiences.

First, we asked for the real-time datasets to the SJSU parking manager. He provided some useful documents (except private data), which helped us a lot with ideas about the project. We started the project with identifying entities, relationships and designing an ER diagram; we forward-engineered to create a smaller replica of SJSU parking database based on information provided on the SJSU parking website. For this project, we downloaded the required fictitious data from mockaroo.com. Through this project, we learned to create a database and how to analyze the database using various queries we were taught in class. We were able to find insights from the database to make the parking system more efficient by using SQL queries. We ran meaningful queries on the database to get the impactful insights.

The objectives are a detailed study of parking infrastructure, permit types and operations; identify entities in the parking database and establish relations among the entities; create an ER diagram, and generated a database to perform meaningful SQL queries. The database we created meets the expectations of all the required entities for performing a variety of queries based on user preference.

During this project, we faced some challenges: first was the unavailability of SJSU Parking Services real-time dataset for users. In this case, we created similar database our own, which means some dummy data do not comply with SJSU’s Parking policies. The second challenge we faced was to agree on the same idea about the table’s normalization and to set foreign/primary keys for an ER diagram.

The future scopes for this project are: Work on SJSU real-time parking datasets; perform and analyze the machine learning algorithms for the real dataset; work towards a smart parking system; explore alternative modes of commute for effective parking management system using real-time dataset.

In the end, we would like to express our sincere gratitude to Prof. Ashraf I Shirani, who has supported us through his patience and knowledge. We are also grateful to him for teaching us and playing a key role in generating interest in this field and providing us with the guidance and motivation whenever needed. We believe we were able to fulfill most of our objectives and we did learn a lot throughout the project.

# References

*About SJSU*. (n.d.). Retrieved November 2019, from http://www.sjsu.edu/about/

*Additional Parking*. (2019, September 9). Retrieved November 2019, from

http://www.sjsu.edu/parking/maps/additional/

Choi, J. (2019, November 8). *Customer Citation Data Schema*. Retrieved November 2019, from

https://mockaroo.com/b14a6070

Choi, J. (2019, November 8). *Occupancy Data Schema*. Retrieved November 2019, from

https://mockaroo.com/c91c2860

*Citation Fines*. (2013, July 10). Retrieved November 2019, from

http://www.sjsu.edu/parking/citations/fines/

*Employee Permits*. (2019, August 1). Retrieved November 2019, from

http://www.sjsu.edu/parking/permits/employee/

*Housing Virtual Permits*. (2019, July 31). Retrieved 2019 November, from

http://www.sjsu.edu/parking/permits/housing/

Koouri, P. (2019, November 8). *Customer Citation Data Schema*. Retrieved November 2019, from

https://mockaroo.com/b14a6070

*Master Plan 2001*. (2001, October). Retrieved November 2019, from

http://www.sjsu.edu/fdo/departments/pdc/whatwedo/masterplan/

*Motorcycle Permits*. (2013, July 10). Retrieved November 2019, from

http://www.sjsu.edu/parking/permits/motorcycle/

*Park & Ride Permits*. (2013, July 10). Retrieved November 2019, from

http://www.sjsu.edu/parking/permits/park\_n\_ride/

*Parking Garages*. (2019, September 5). Retrieved November 2019, from

http://www.sjsu.edu/parking/maps/garages/

Parking Information. (2019, November 14). Retrieved November 3, 2019, from

<http://www.sjsu.edu/hsr/facultyhandbook/>.

*Parking Lots*. (2019, September 9). Retrieved November 2019, from

http://www.sjsu.edu/parking/maps/lots/

*Student Commuter Virtual Permits*. (2019, September 18). Retrieved November 2019, from

http://www.sjsu.edu/parking/permits/student/

*Visitor / Guest Parking*. (2019, October 4). Retrieved November 2019, from

http://www.sjsu.edu/parking/permits/visitor/daily\_short-term/index.html