# Dimensional Modeling and Analysis for Car Repair Shop

Part 2: Data Ingestion, Preprocessing, Analysis and actionable Recommendations

### 1.0 <u>Introduction</u>

The second part of this project involved setting up a relational database and importing data from provided CSV files to facilitate comprehensive analysis of the car repair shop's operations. The data includes customer information, vehicle details, invoices, job descriptions, and parts usage. The goal was to ensure that the data was clean, properly formatted, and correctly loaded into the database, setting the stage for detailed analysis.

#### **Objectives**

### 1. Data Ingestion and Preparation:

- o Import the provided CSV files into a relational database.
- Ensure that the data is clean, properly formatted, and indexed for efficient querying.

#### 2. Data Analysis Tasks:

- Perform customer analysis, vehicle analysis, job performance analysis, parts usage analysis, and financial analysis.
- **Data Consistency Checks:** The data was carefully examined for inconsistencies, errors, or missing values. For instance, discrepancies in job IDs and part associations were identified and corrected.

## 2.0 Data Acquisition and Preprocessing

#### 1. Data Source:

The raw data for this analysis was sourced from <u>CSV files</u> which were carefully inspected to understand the structure and contents of the data and prepared for import into the relational database. The data included customer information, vehicle details, invoices, job descriptions, and parts usage.

### 2. Data Cleaning and Validation:

To ensure data accuracy and consistency, a number of data cleaning steps were performed:

- Address Standardization: The customer address field was split into separate fields (Street, City, State, Zip) for better data organization and easier querying for analysis.
- Vehicle Data: The registration number was removed as it was considered irrelevant for the analysis. The vehicle data was then formatted appropriately.
- Invoice Data: The date column was formatted to ensure consistency in representing years, months, and days.

#### 3. Data Import:

The prepared and formatted CSV files were imported into the corresponding tables in the relational database using SQL LOAD DATA INFILE statements. The tables included CustomerDim, VehicleDim, InvoiceDim, JobDim, PartsDim, and LocationDim.

```
165
       -- Load Customer data
166 •
       LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/customer.csv'
167
       INTO TABLE CustomerDim
       FIELDS TERMINATED BY ','
169
       ENCLOSED BY ""'
      LINES TERMINATED BY '\n'
170
      IGNORE 1 ROWS
171
172
       (CustomerID, CustomerName, Street, City, State, Zip, Phone);
173 • Select * From CustomerDim;
174
175
        -- Load Vehicle data
176 • LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/vehicle.csv'
       INTO TABLE VehicleDim
177
178
       FIELDS TERMINATED BY ',
179
       ENCLOSED BY ""
       LINES TERMINATED BY '\n'
       IGNORE 1 ROWS
181
182
       (VehicleID, Make, Model, Year, Color, VIN, Mileage, CustomerID);
183 • Select * From VehicleDim:
184
185
        -- Load Invoice data
186 • LOAD DATA INFILE 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/invoice.csv'
187
      INTO TABLE InvoiceDim
188
      FIELDS TERMINATED BY ','
      ENCLOSED BY ""
189
       LINES TERMINATED BY '\n'
190
191
       IGNORE 1 ROWS
       (InvoiceID, InvoiceDate, Subtotal, SalesTaxRate, SalesTax, TotalLabour, TotalParts, Total, CustomerID, VehicleID);
```

Fig 1. Screenshots of load data infile statements for CustomerDim, VehicleDim and InvoiceDim

 Location Data: Since there was only one shop location, it was directly inserted into the database using an SQL statement.

```
136 • INSERT INTO LocationDim (LocationID, ShopName, Street, City, State, Zip, Phone)
VALUES (1, 'Latino Garage', '111McPhillips', 'Winnipeg', 'MB', 'R3J 1X7', '204-984-8458');
```

Fig 2. Screenshots of inserting data into LocationDim

#### 4. Data Validation:

 After loading the data, each table was reviewed to ensure successful data import. This included verifying the correctness and completeness of the data in each table to ensure data integrity and accuracy. Example shared below.

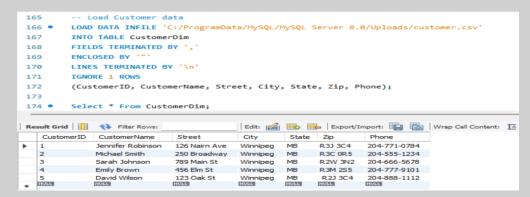


Fig 3. Results for reviewing CustomerDim

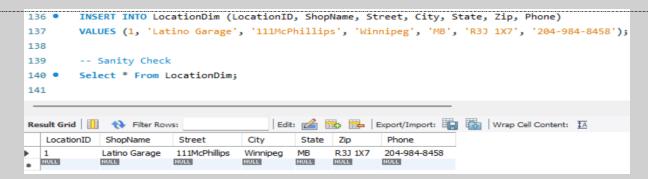


Fig 4. Results for reviewing LocationDim

### 5. Data Cleaning:

#### **Error Correction and Validation:**

During the review, it was noticed that Job IDs 1 and 4 were not represented and Job ID 3 was wrongly placed in the PartsDim table, leading to discrepancies.

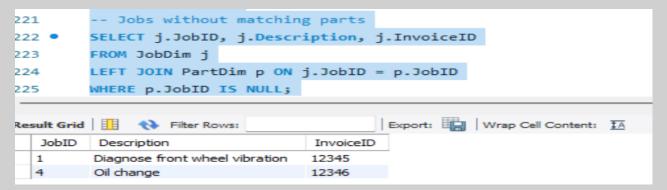


Fig 5. Results for JobID without Matching PartID

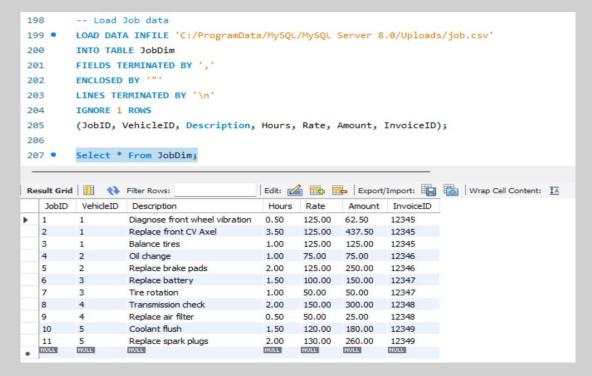


Fig 6. Results for reviewing JobDim

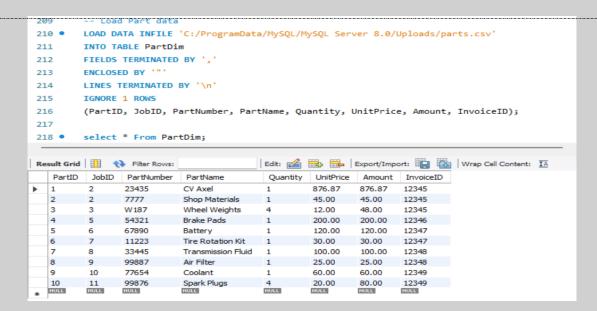


Fig 7. Results for reviewing PartDim

### **Correcting Data Associations:**

SQL queries were written to update the incorrect associations of JobID 1 and 3 and also inserted data for JobID 4 to avoid losing data.

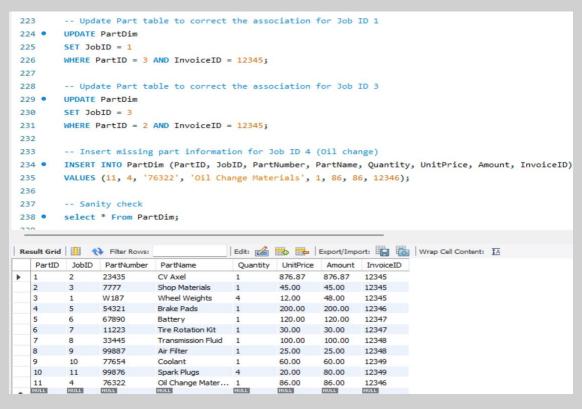


Fig 8. Updating, inserting missing values and validating check for PartDim.

#### **Sanity Checks**

After making corrections, sanity checks were performed to ensure the data was correctly loaded and accurately represented in the database.

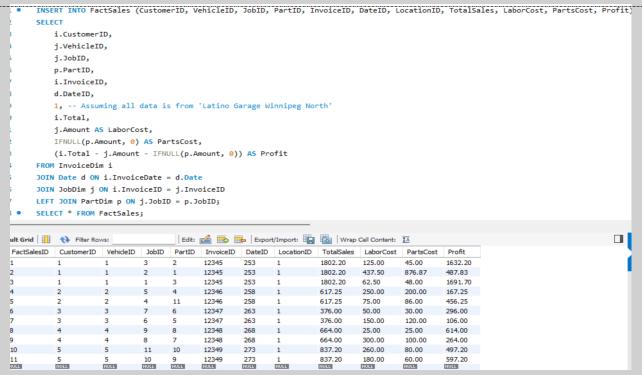


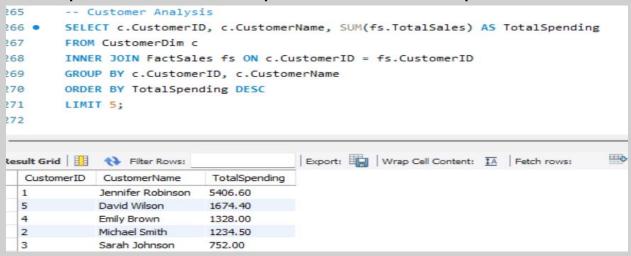
Fig 8. Validating FactSales.

### 3.0 Data Analysis

### Customer Analysis

The customer analysis aimed to provide insights into customer spending behavior, frequency of visits, and spending patterns. The following SQL queries were used to derive the necessary information:

#### 1. Top 5 Customers Who Have Spent the Most on Vehicle Repairs and Parts



By identifying the top-spending customers, the analysis highlights key contributors to overall revenue. These insights can inform targeted loyalty programs, personalized marketing campaigns, and customer retention strategies. Understanding customer spending patterns is crucial for optimizing resource allocation and maximizing customer lifetime value.

### 2. Average Spending of Customers on Repairs and Parts

```
272
273 • SELECT AVG(TotalSpending) AS AverageSpendingPerCustomer

274 FROM (
275 SELECT c.CustomerID, SUM(fs.TotalSales) AS TotalSpending

276 FROM CustomerDim c
277 INNER JOIN FactSales fs ON c.CustomerID = fs.CustomerID

278 GROUP BY c.CustomerID

279 AS CustomerSpending;

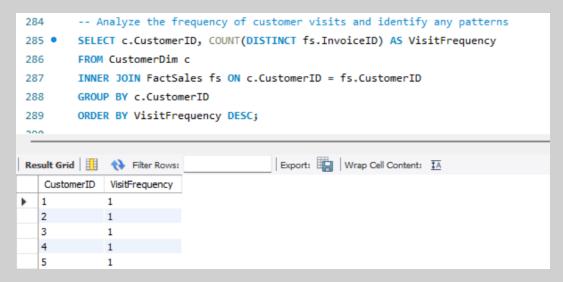
Result Grid Filter Rows:

AverageSpendingPerCustomer

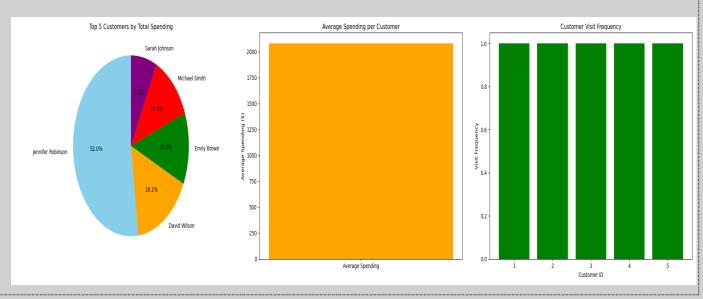
2079,100000
```

The average spending per customer was calculated to understand the general spending behavior of the customer base. This metric is useful for benchmarking individual customer spending against the overall average.

## 3. Frequency of Customer Visits and Identification of Patterns



The frequency of customer visits was analyzed to identify patterns in customer engagement. Understanding visit frequency can help in designing retention strategies and improving customer service.



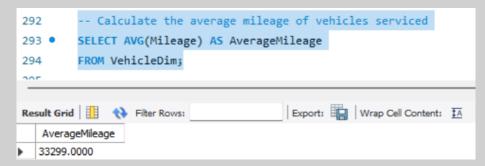
The pie chart highlights the significant contribution of **Jennifer Robinson** to the total revenue, indicating she is a high-value customer. Targeted marketing and personalized services could be beneficial for retaining such valuable customers. The average spending per customer is about **\$2,000**. Each of the top 5 customers has a visit frequency close to 1. High visit frequency by top spenders suggests strong customer loyalty and satisfaction. Maintaining or improving service quality is crucial to ensure these customers continue to return

Understanding customer behavior is essential for enhancing customer satisfaction and loyalty.

### Vehicle Analysis

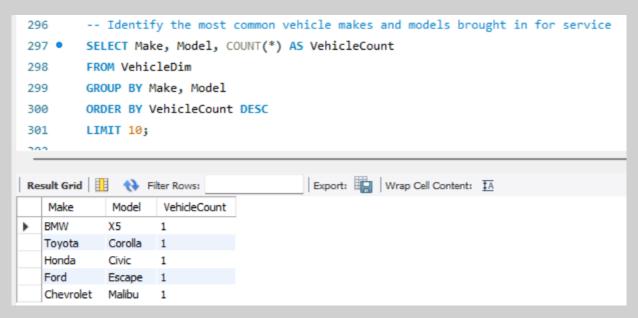
The vehicle analysis focuses on understanding the characteristics and trends related to the vehicles serviced at the car repair shop. The following SQL queries were used to derive the necessary information:

### 1. Calculate the Average Mileage of Vehicles Serviced



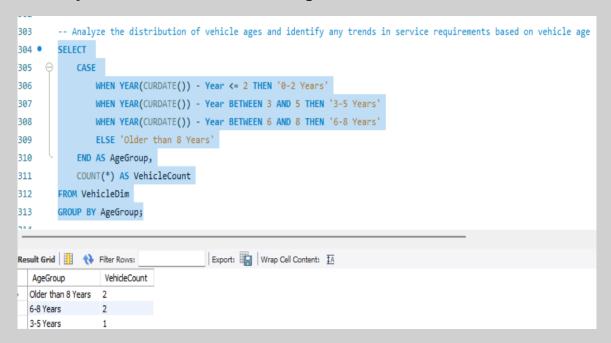
The query calculated the average mileage of all vehicles serviced, providing insight into the typical wear and tear of vehicles that visit the repair shop.

### 2. Identify the Most Common Vehicle Makes and Models Brought in for Service



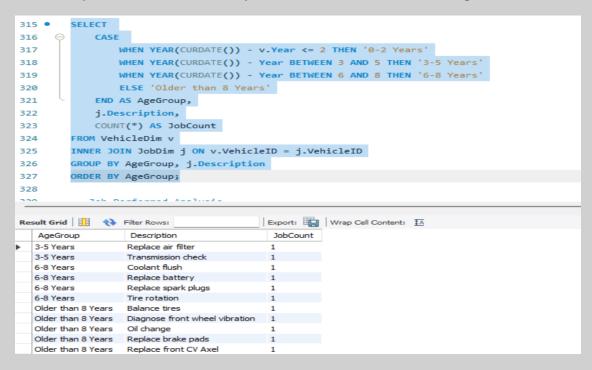
The query identified the most frequently serviced vehicle makes and models, helping the shop understand which types of vehicles are most commonly brought in for repairs and potential service specializations.

### 3. Analyze the Distribution of Vehicle Ages

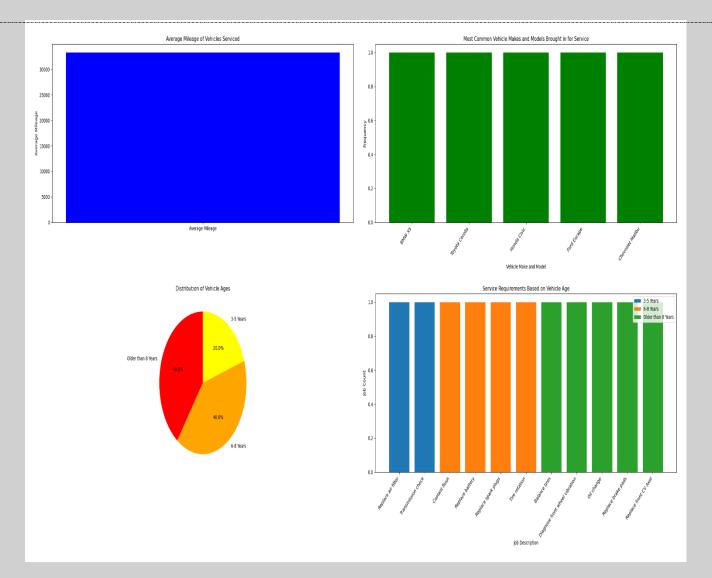


By categorizing vehicles into age groups, the analysis revealed the proportion of different vehicle age segments serviced by the shop.

### 4. Identify Trends in Service Requirements Based on Vehicle Age



The analysis correlated vehicle age with the types of services required, identifying potential service patterns based on vehicle lifecycle. This helps understand what types of services are needed more frequently for different age groups.



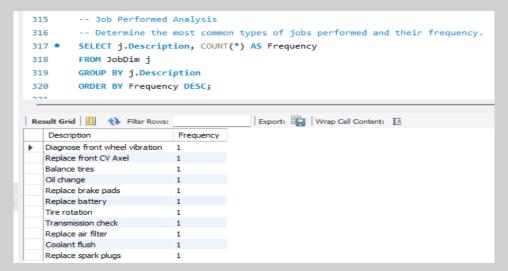
The average mileage of serviced vehicles is about 30,000 miles. Each of these models has a similar frequency of service. Identifying the most common vehicle makes and models helps the shop stock the right parts and anticipate common issues associated with these vehicles. 40% of the serviced vehicles are older than 8 years. Another 40% of the serviced vehicles are between 6-8 years old. Then 20% of the serviced vehicles are between 3-5 years old. Understanding the age distribution of vehicles helps the shop anticipate the types of services required for older vehicles, which may need more extensive maintenance. Older vehicles (6-8 years and older than 8 years) frequently require more intensive services like replacing spark plugs, brake pads, and CV axles. By understanding service needs based on vehicle age, the shop can better prepare for common repairs and maintenance tasks, ensuring they have the necessary parts and expertise.

These insights can inform inventory management, service planning, and customer engagement strategies.

### 👃 Job Performance Analysis

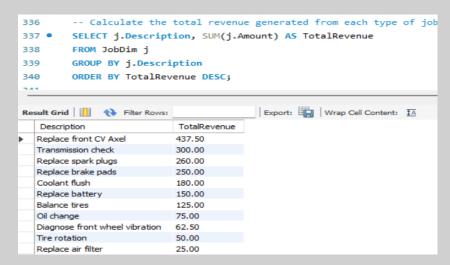
The job performance analysis focuses on understanding the types of jobs performed, their frequency, the revenue generated from each type of job, and the average costs associated with different jobs. The following SQL queries were used to derive the necessary information:

### 1. Determine the Most Common Types of Jobs Performed and Their Frequency



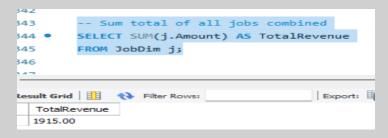
The query identified the most frequently performed jobs, helping the shop understand which services are in high demand.

### 2. Calculate the Total Revenue Generated from Each Type of Job



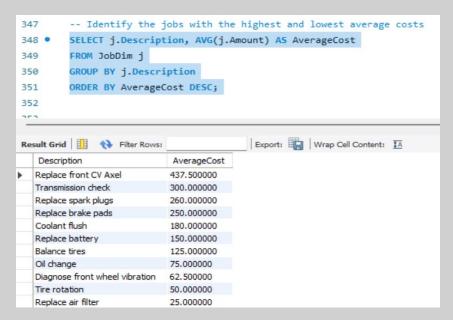
The query calculated the total revenue generated from each type of job, providing insight into the financial contribution of different services.

#### 3. Calculate the Sum Total of All Jobs Combined

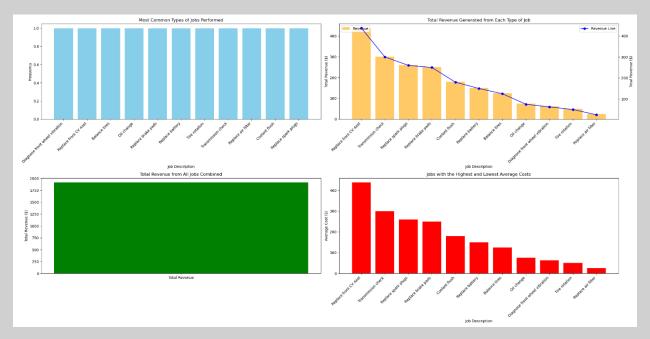


The query calculated the total revenue generated from all jobs combined, giving an overall picture of the financial performance.

### 4. Identify the Jobs with the Highest and Lowest Average Costs:



The query identified the jobs with the highest and lowest average costs, providing insight into the cost dynamics of different services.



Jobs such as diagnosing front wheel vibration, replacing the front CV axle, balancing tires, oil changes, and replacing brake pads are frequently performed. Understanding the most common jobs helps the repair shop allocate resources efficiently and ensure they have the necessary parts and expertise for these frequently performed tasks. Replacing the front CV axle generates the highest revenue, followed by transmission checks and replacing spark plugs. Then, the revenue decreases gradually for other job types like oil changes, tire rotation, and air filter replacement. Lower cost jobs include oil changes, tire rotation, and replacing air filters. Identifying which jobs generate the most revenue helps the shop focus on high-value services and optimize pricing strategies.

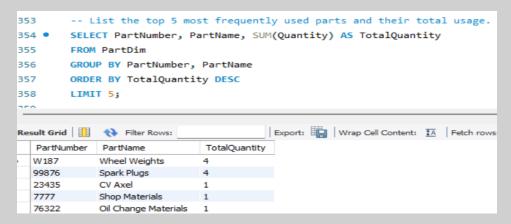
The total revenue from all jobs combined is about \$1,900. The overall revenue figure helps the shop assess its financial performance and set revenue targets for future periods. Understanding the cost structure of different jobs helps the shop in pricing services appropriately and identifying areas for cost optimization.

This information can be used to optimize service offerings, pricing strategies, and resource allocation.

## Part Usage Analysis

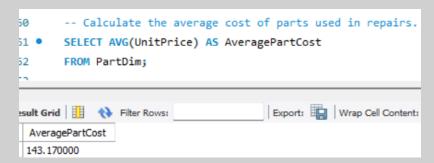
The part usage analysis focuses on understanding the usage and financial aspects of parts used in vehicle repairs. The following SQL queries were used to derive the necessary information:

#### 1. List the Top 5 Most Frequently Used Parts and Their Total Usage



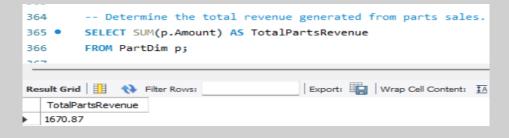
The top 5 most frequently used parts, providing insight into which parts are in high demand and need to be stocked regularly.

#### 2. Calculate the Average Cost of Parts Used in Repairs:

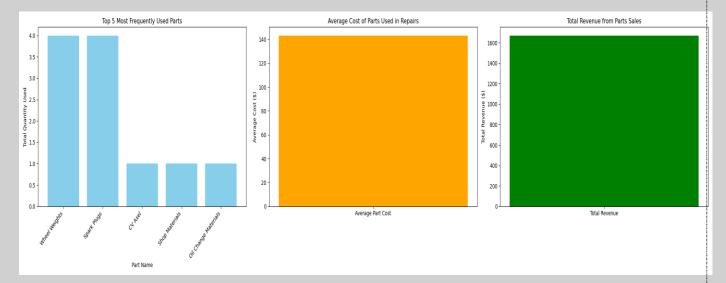


The calculated average cost of parts used in repairs, provide insight into the typical cost associated with part replacements and repairs.

#### 3. Determine the Total Revenue Generated from Parts Sales



The total revenue generated from parts sales, gives an overall picture of the financial performance related to parts.



Wheel Weights and Spark Plugs are the most frequently used parts, each with a total quantity of around 4. Other frequently used parts include CV Axel, Shop Materials, and Oil Change Materials. Knowing which parts are used most frequently helps the shop maintain adequate inventory levels to meet demand and avoid service delays due to part shortages. The average cost of parts used in repairs is around \$140. The total revenue from parts sales is around \$1,600. This overall revenue figure highlights the importance of parts sales as a significant revenue stream for the repair shop. Strategies to boost parts sales, such as bundling services with parts or offering promotions, could be beneficial.

## Financial Analysis

The financial analysis focuses on understanding the revenue generated from labor and parts, the overall profitability of the repair shop, and the impact of sales tax on total revenue. The following SQL queries were used to derive the necessary information:

#### 1. Calculate the Total Revenue Generated from Labor and Parts for Each Month



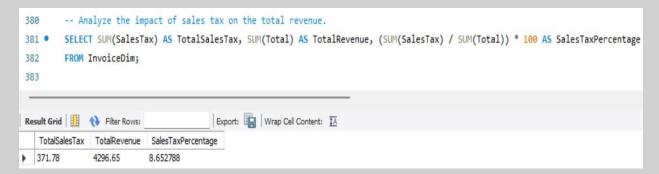
The total revenue generated from labor and parts for each month, providing insight into the monthly financial performance.

### 2. Determine the Overall Profitability of the Repair Shop:

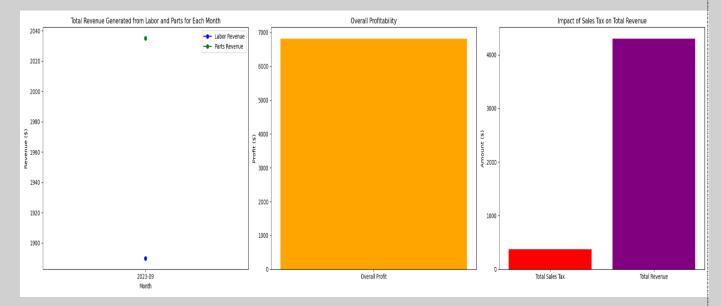


The query calculated the overall profitability of the repair shop, giving an overall picture of the financial health of the business.

### 3. Analyze the Impact of Sales Tax on the Total Revenue:



The query analyzed the impact of sales tax on total revenue, providing insight into the percentage of revenue contributed by sales tax.



This line plot displays the total revenue generated from labor and parts for September 2023. Labor revenue for September 2023 is almost \$1,900 and Parts revenue for September 2023 is slightly higher at around \$2,040. This information helps the shop understand the contributions of labor and parts to the overall revenue, which is essential for budgeting and financial planning. The overall profit is about \$6,800. Recognizing the overall profitability helps the shop assess its financial health and make informed decisions about investments, cost management, and pricing strategies. Total sales tax (8.65%) collected is relatively small (\$371.78) compared to the total revenue (\$4,296.65). It helps in understanding the impact of tax policies on the business and aids in compliance and financial reporting.

These insights can inform financial planning, pricing strategies, and tax management. By examining revenue, costs, and profitability metrics, the shop can identify areas for improvement, such as cost reduction and revenue enhancement.

## 4.0 Optimization Recommendations

Based on the analysis of the customer, vehicle, job performance, parts usage, and financial data, the following actionable recommendations can help Latino Garage optimize operations and enhance the overall performance of the repair shop:

### **Identifying Underperforming Services**

Jobs such as "Replace air filter" and "Tire rotation" show low revenue and frequency.

Latino Garage needs to evaluate these services to identify reasons for their underperformance. The Garage should consider marketing efforts to promote these services or bundle them with other popular services to increase their uptake.

### **Stock Management for Frequently Used Parts**

Parts like "Wheel Weights" and "Spark Plugs" are used frequently. The Garage should ensure these parts are always in stock to avoid service delays. It should establish a robust inventory management system to track the usage and reorder these parts in a timely manner.

### **Customer Loyalty Programs**

Customers like **Jennifer Robinson** and **David Wilson** are top spenders. It is important for the Garage to implement a loyalty program to reward these customers for their continued business. Offer discounts, priority service, or special promotions to encourage repeat visits and maintain customer loyalty.

### **Scheduling Adjustments for Frequent Jobs**

Jobs such as "Diagnose front wheel vibration" and "Replace front CV axle" are common. Adjust scheduling and staffing to ensure these frequent jobs can be handled efficiently. Consider cross-training staff to handle these jobs to reduce wait times and improve customer satisfaction.

#### **Dynamic Pricing Strategy**

Jobs like **"Replace front CV axle"** generate significant revenue. Implement a dynamic pricing strategy where prices for high-value and high-demand jobs can be adjusted based on market demand, customer willingness to pay, and competition.

### **Enhanced Training Programs**

Latino Garage should ensure staff are well-trained to handle a variety of jobs, especially those that are frequently performed. The Garage can develop continuous training programs to enhance the skills of the staff. This will improve service quality and efficiency.

#### **Customer Feedback Mechanism**

Gather feedback from customers to understand their needs and expectations better. Implement a structured feedback mechanism to collect customer opinions regularly. Use this feedback to make data-driven improvements in services and customer interactions.

#### **Digital Transformation**

Improve the customer experience by integrating digital tools. Implement an online booking system and a mobile app where customers can schedule appointments, track service progress, and receive notifications about service completion and promotions.

### **Promotional Campaigns for Low-Traffic Periods**

Identify periods with low customer traffic and offer promotions. Develop targeted promotional campaigns and discounts during slow periods to attract more customers and even out the workload throughout the month.

#### **Green Initiatives**

Promote eco-friendly services and parts. Introduce green initiatives such as using environmentally friendly parts and offering services that reduce environmental impact. Market these initiatives to attract environmentally conscious customers.

### 5.0 Conclusion

The comprehensive analysis of the car repair shop's operations, including customer behavior, vehicle trends, job performance, parts usage, and financial metrics, has provided valuable insights into areas of strength and opportunities for improvement. By implementing the recommendations, Latino Garage can optimize its operations, improve customer satisfaction, and increase profitability

#### **Links to Resources**

- Relational Database Containing the Imported CSV Files: <u>Link to Database</u>
- SQL Script for Creating Tables and Importing Data: Link to SQL Script
- Notebook for Visualizations on the analysis: <u>Link to Notebook</u>