

Report

# **IT-300 Business Intelligence and Database Management Systems**

Business Intelligence Research Car Renting & Customer Ratings Data

In the US

Authors: Submitted to:

Aya Ghannouchi Prof. Manel Abdelkader

Ali Abdelmouleh Prof. Ameni Azzouz

Karim Ben Amar

Sarra Azaiez

Contents :

1 -Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2 -Implementation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.1 Data Gathering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.2 Data Preparation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2.3 Data Storage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. . . 6

2.3.1 Storage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2.3.2 Fact . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.3.3 Dimensions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2.4 Data Visualization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 10

3- Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .. 12

# 1-Introduction:

This business intelligence project is dedicated to analyzing car rental and customer rating data. The objective is to explore the impact of several factors, including customer age, gender, car categories, car models, and types of cars rented, on customer ratings.

By uncovering these correlations, we can derive valuable insights into customer behavior and preferences in the car rental industry. The goal is to assist car rental companies in enhancing their services, thus leading to increased customer satisfaction and the development of more effective strategies for the industry. Through this project, we aspire to contribute to elevating the overall customer experience within the car rental sector.

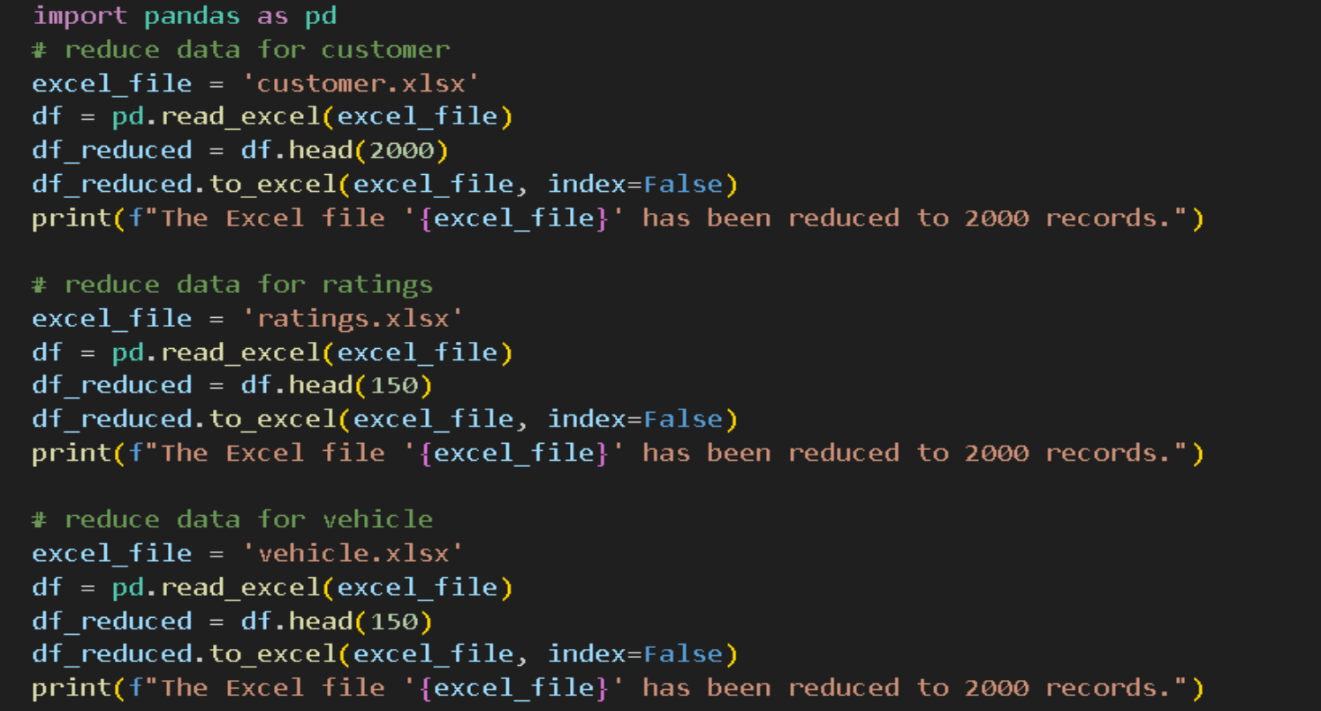
# 2 -Implementation:

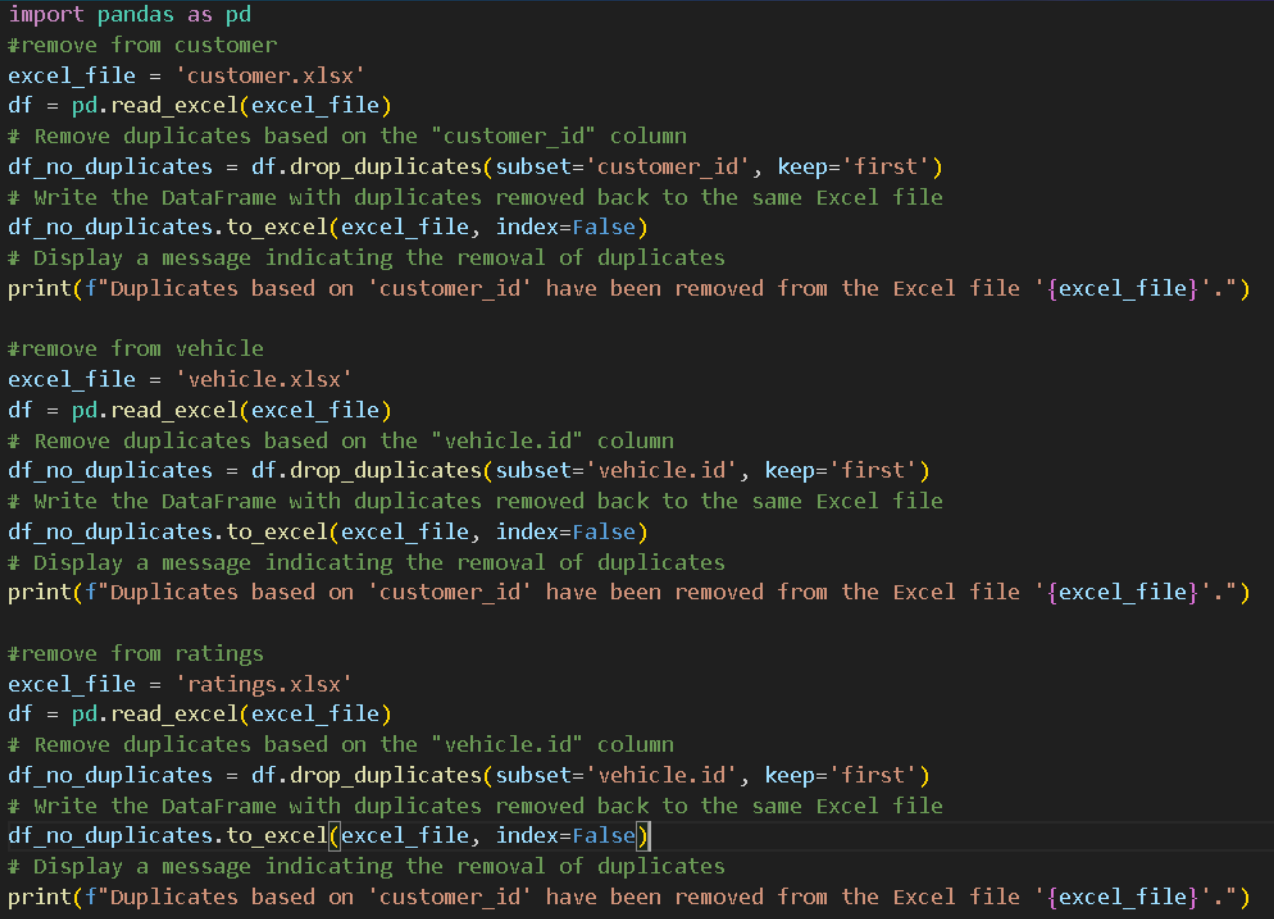
2.1 Data Gathering:

For this project, we gathered the dataset on car rentals and customer ratings in the USA from Kaggle. The dataset includes a comprehensive range of information, including details on various car rental transactions and corresponding customer ratings. To enhance our analytical capabilities, we generated additional columns through a combination of Python programming and Excel spreadsheet functionalities. These supplementary columns provide further insights into the data, allowing us to conduct a more comprehensive analysis. The dataset link on Kaggle is provided for reference.

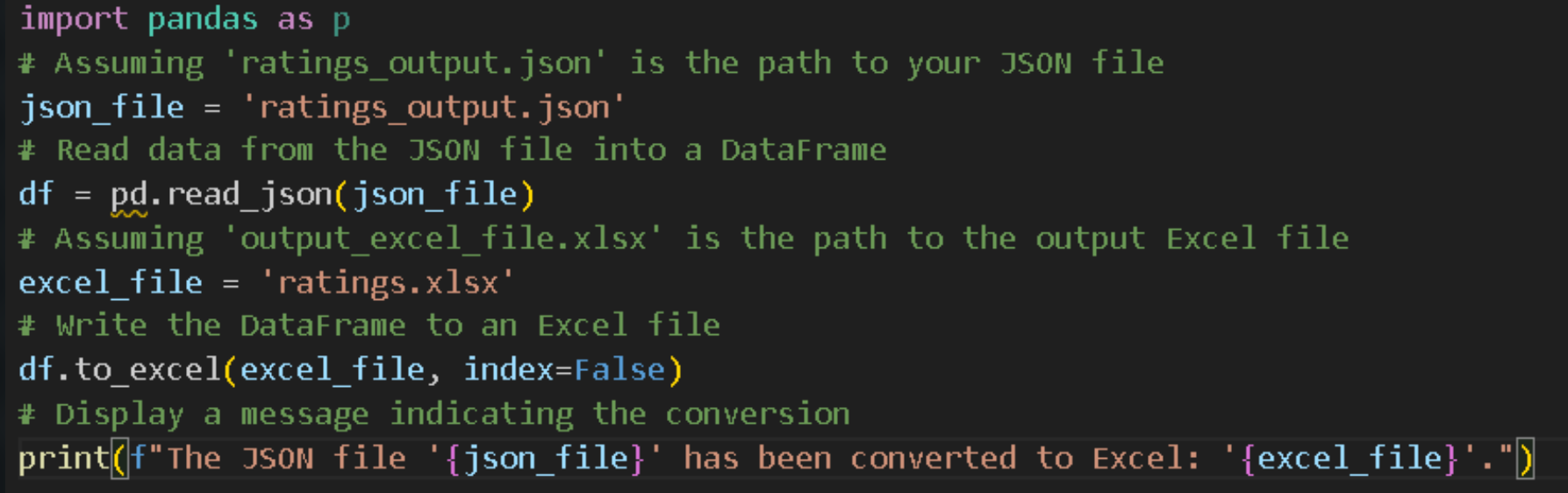
2.2 Data Preparation:

In the data preparation phase (ETL), Python was employed. The original dataset presented issues with duplicates and an extensive volume of data. Using the provided code, we effectively removed duplicates and opted for the first 2000 customers, 150 initial vehicles, and 150 initial ratings to streamline subsequent analysis.

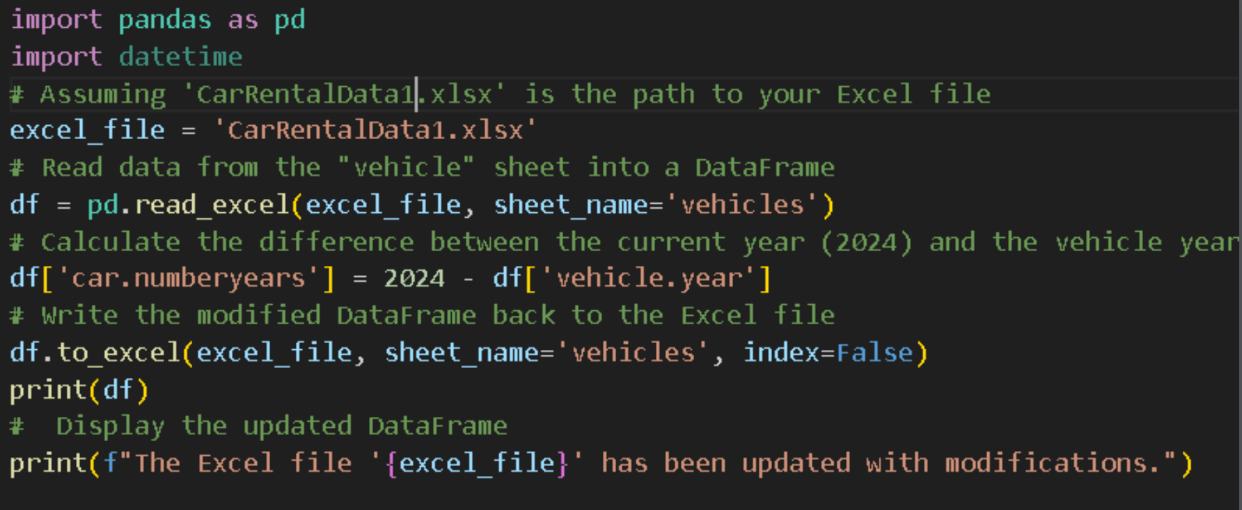




We noticed the ratings data was in JSON format. To work smoothly with Power BI and avoid issues, we converted it to XLSX format.



In the vehicle data, we wanted to determine the age of each vehicle. By subtracting the manufacturing year from 2024 using the provided code, we successfully calculated the vehicle age.



To identify the optimal season for the business, we introduced a new column named "season" in the customer data. This column was populated based on the month in which the customer rented the vehicle.



2.3 Data Storage:

2.3.1 Storage:

For data storage, we used PgAdmin to map tables in our code and perform the star schema and PostgreSQL as the database management system.

To facilitate seamless communication between the two, we used the ERD per table tool in Postgre SQL.

We have three main tables in the database, Customers, Vehicles, and Ratings.

Attached here is the code for the Postgre SQL tables:

* Customers: Represents the Customer information.

A screenshot of a computer

Description automatically generated

* Vehicles: Represents the Vehicle information: brand, model, type, category, fuel type, and year of make.

A screenshot of a computer code

Description automatically generated

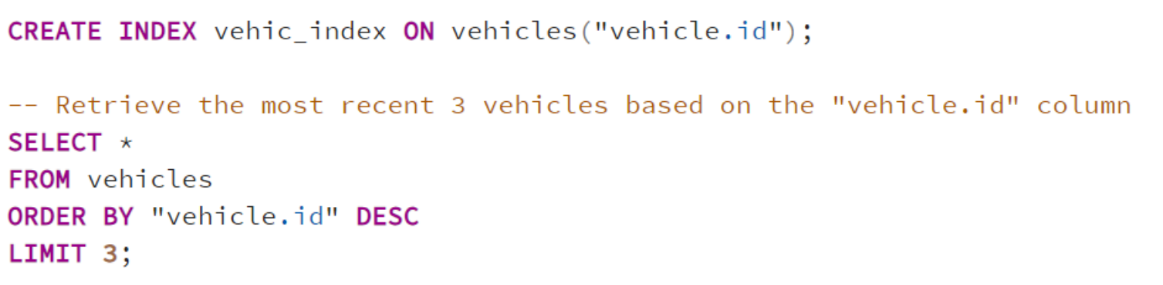
* Ratings: Represents the car rating, customer review of that specific vehicles, trips, and reviews for every vehicle, and finally the total of rented days and the price per rent per day. to evaluate the overall renting experience.

A computer code with text

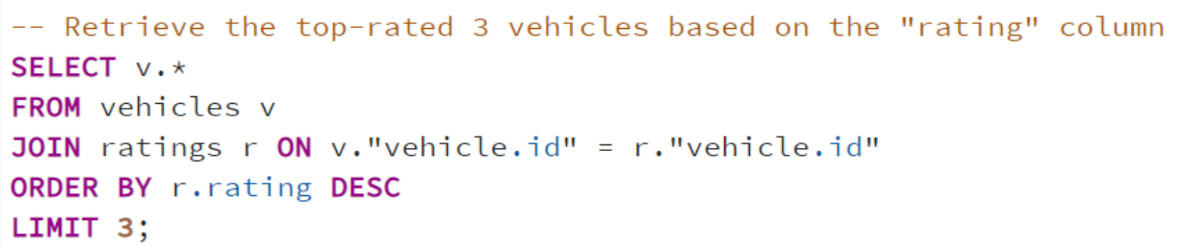
Description automatically generated

These are four additional examples of queries that we can run in PgAdmin using the indexing and aggregation tools to display useful output:

* Most recent three vehicles in the database:



* Top rated three vehicles:



* Average rating for each vehicle:

A close-up of text

Description automatically generated

* Top rented days for each vehicle

A computer screen shot of a vehicle

Description automatically generated

A screenshot of a computer

Description automatically generated

Figure 1: Data Warehouse Schema

2.3.2 Fact:

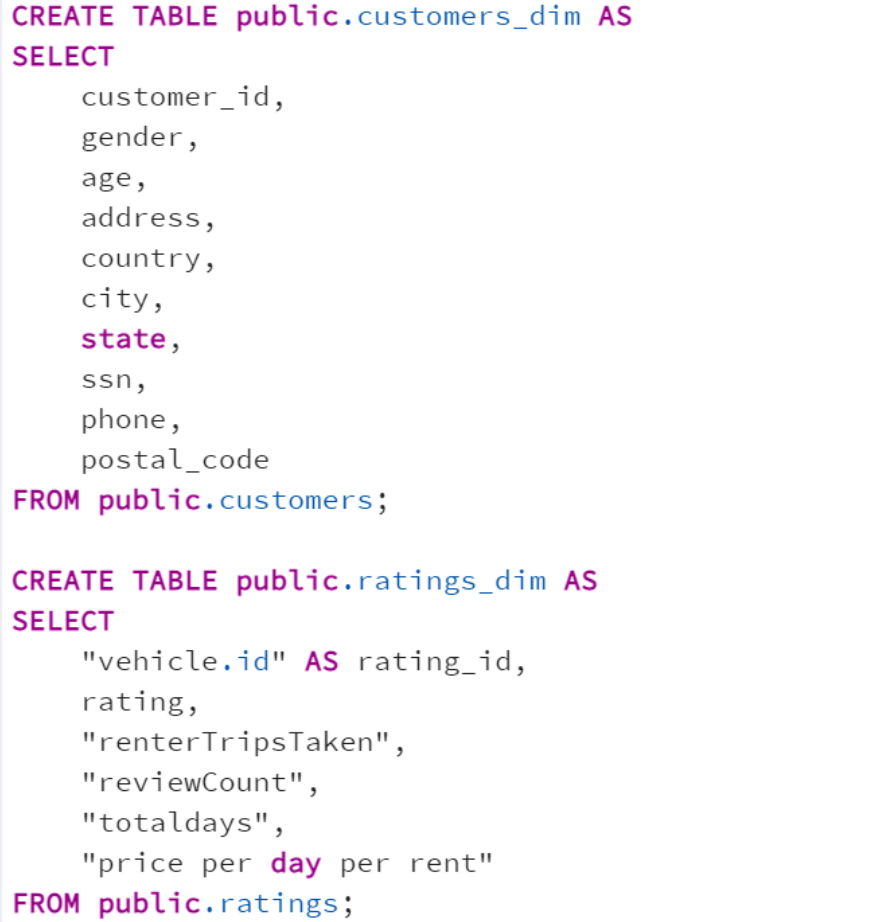
The fact table includes details about the vehicle, customer, and rental information associated with the renting event.

* Rental date
* Rental time
* Rental duration

2.3.3 Dimensions:

Dimensions included in this data are:

* Customers
* Vehicles
* Ratings

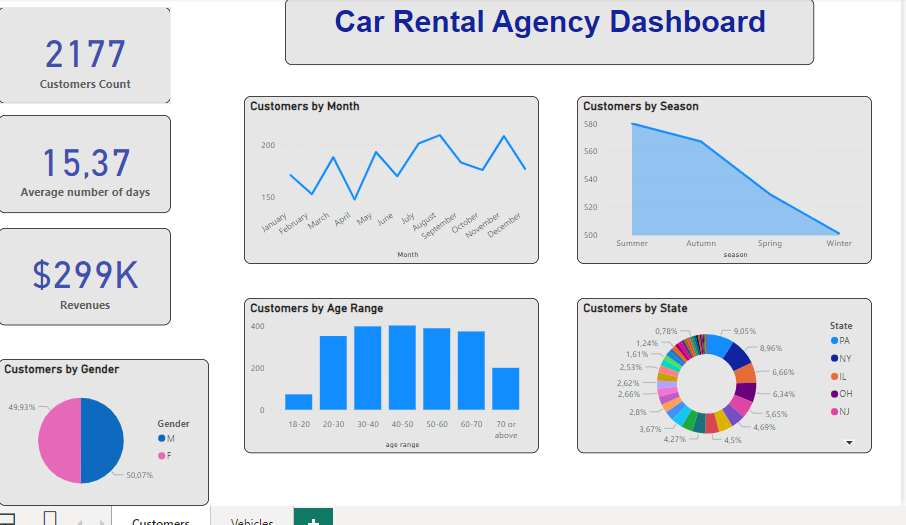


A screenshot of a computer code

Description automatically generated

* Our data warehouse follows a star schema, where the main "rental" table is surrounded by supporting tables: "customers," "vehicles," and "ratings." This setup simplifies data analysis, making it easier to understand and retrieve information about customer behaviors, vehicle details, and ratings.

2.4 Data Visualization:



A screenshot of a computer

Description automatically generated

1. The most common age range for customers is between 40 and 50 years.

2. Peak customer activity occurs in July and August.

3. The majority of customers prefer renting cars during the summer season.

4. The customer distribution across multiple states includes Pennsylvania (PA) at 9.0%, New York (NY) at 8.96%, Illinois (IL) at 6.66%, and Ohio (OH) at 6.34%. These percentages reflect a varied customer presence across these states.

5. The customer base is fairly balanced in terms of gender, with approximately 50.07% identifying as male and 49.93% as female.

6. The average customer rating for rented vehicles is 4.92.

7. Gasoline is the most commonly available vehicle fuel type.

8. The top three vehicle companies with the highest days rented are BMW, Ford, and Chevrolet.

9. The top three revenue-generating vehicle companies are Tesla, Chevrolet, and BMW.

10. Cars are the most frequently available vehicle type.

11. High-end vehicles are the most popular category, followed by medium-range vehicles.

12. Number of Reviews: We have received feedback through 1171 reviews, indicating active customer engagement.

13. Number of Rentals: A total of 2177 rentals have been processed, showing an elevated level of customer activity.

14. Total Number of Days Rented: Customers have rented vehicles for a combined total of 16,735 days, illustrating substantial usage of the rental service.

15. Customer Count: There are 2177 unique customers, representing a diverse customer base.

16. Average Number of Days: On average, customers rent vehicles for 15.37 days, giving an idea of the typical rental duration.

17. Revenue: The total revenue generated from these rentals is $299,000, reflecting the financial success of the car rental business.

3 -Conclusion:

The car rental data reveals peak activity in July-August, with a balanced age and gender distribution. Pennsylvania (PA) stands out in customer count. High-end cars, led by BMW, Ford, and Chevrolet, drive revenue. Maintaining a high 4.92 rating, the agency can strategically target advertising and enhance services for continued success in the USA.