Milestone 1 Report: Business Level Analysis

Introduction

The objective of Milestone 1 is to analyze businesses in Arizona (AZ) based on attributes such as reviews, ratings, service types, and locations. This analysis focuses on identifying patterns and insights to assess performance and customer preferences in specific business categories. The data has been processed from JSON to Parquet for efficient querying and filtered using Spark SQL.

Methods and Analysis

1. Rating Distribution by City (Top 5 Cities)

Objective: Compare average ratings of businesses in the top-performing cities to determine city-wise performance.
Method: Filtered business data for Arizona and grouped by city. Average ratings were calculated using Spark SQL and visualized through a bar graph.
Result: Businesses in Oro Valley received the highest average ratings compared to Tucson, which also performed well.
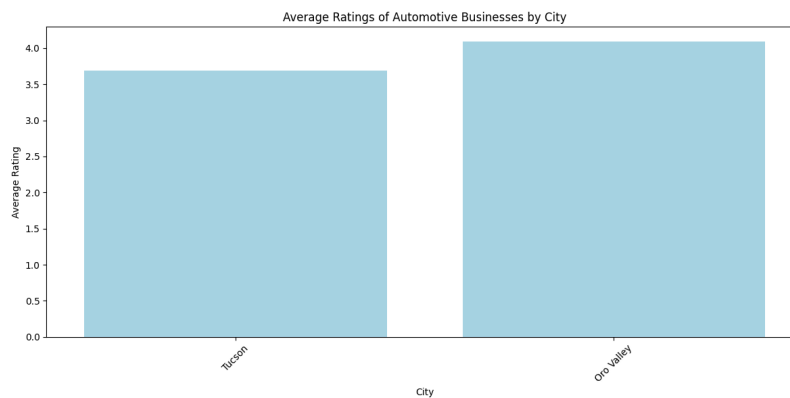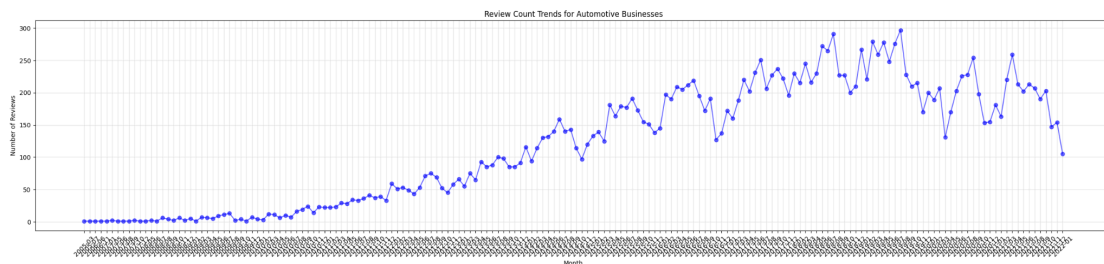

2. Review Count Trends Over Time

Objective: Examine the trend of customer reviews over time to gauge engagement and popularity.
Method: Aggregated monthly review counts from the dataset and visualized trends using a line chart.
Result: A steady increase in review counts over time indicates growing customer interaction, peaking at specific intervals.


3. Rating Distribution by Service Type
Objective: Identify which service types are most favored by customers based on average ratings.
Method: Filtered data by service categories and calculated the average rating for each type. Results were visualized in a bar graph.
Result: Home Window Tinting and Transmission Repair services received the highest average ratings, indicating customer satisfaction in these areas.

Average Ratings by Automotive Service Type
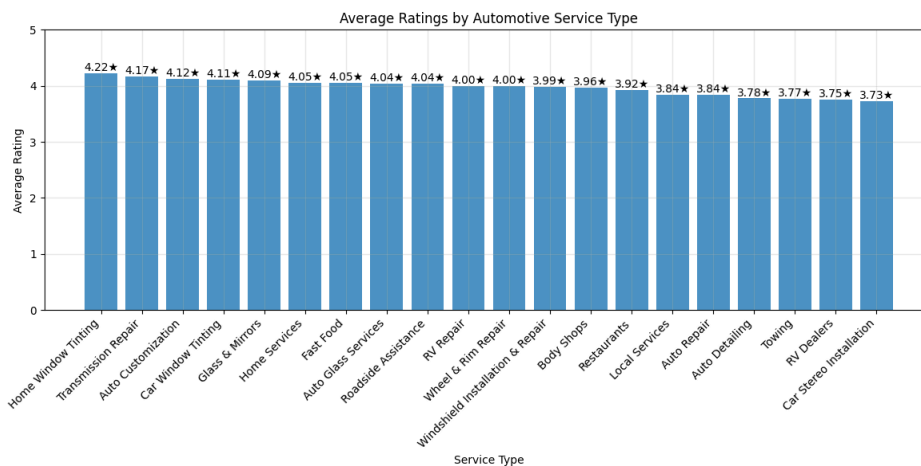
4. Review Sentiment Over Time

Objective: Track positive and negative review sentiment trends to assess customer satisfaction.

Method: Sentiment analysis was conducted on review text, categorizing reviews as positive or negative. The ratio of sentiments was visualized in a stacked area chart.

Result: Positive reviews consistently outweighed negative ones, suggesting overall customer satisfaction.

Review Sentiment Ratio Over Time
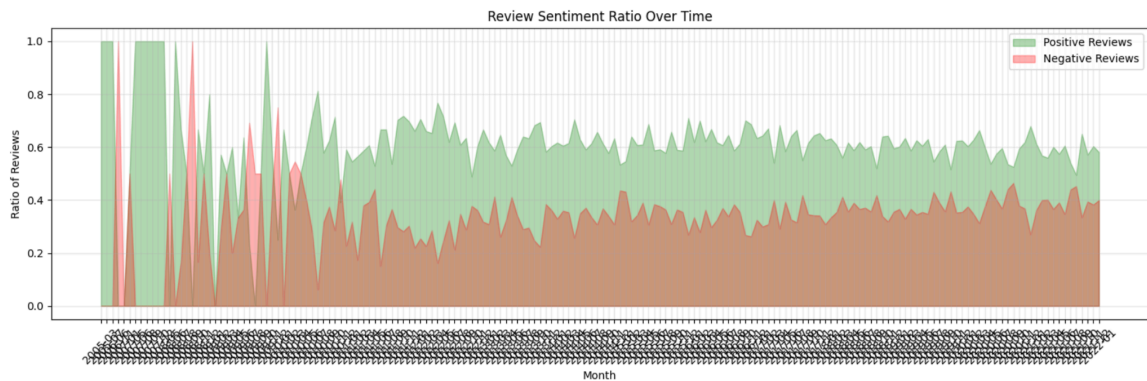
5. Business Hours Analysis

Objective: Explore how opening hours correlate with the number of businesses and their average ratings.

Method: Analyzed the distribution of businesses and their ratings based on opening hours using Spark SQL and visualized the results in bar and line charts.

Result: Most businesses open at 12 PM achieved the highest average ratings, though the majority operate between 7 AM and 9 AM.

Average Rating by Opening Hour