

Online Car Sale Analysis – Final Report

Group Members

- Alex Alvarez
- Drish Sethi]

GitHub Repository

<https://github.com/mceballos92/Online-Car-Sale-Analysis>

Project Summary

This project analyzes used car listings from a popular online marketplace. The goal was to identify pricing trends based on various attributes such as mileage, fuel type, and dealer listing patterns.

Key Steps and Contributions

- Data Cleaning (branch: CleanData)
 - Cleaned dataset using Jupyter and removed irrelevant or duplicate entries.
- High-Level Questions
 - Q1: What factors influence the most listed cars? (branch: HighQ1)
 - Q2: What's the average price of cars with low mileage? (branch: SecPartHQ)
 - Q3: How does popularity affect pricing? (branch: SecPart2)
- Analysis & Visuals
 - Used Python, pandas, and matplotlib.
 - Multiple Jupyter notebooks were created for each analysis part.
- Collaborative Workflow
 - I worked on all branches.
 - All branches were merged to main via pull requests as required.

Proof of Merge

- Final merged files are available in the main branch of the repository.
- All individual contributions were integrated through GitHub workflows.

Files in Repository

- DS311_Group_Project_Option_4.pdf (project instructions)
- Group project clean up.ipynb (main notebook)
- README.md (project overview)
- All dataset files and merged .ipynb contributions

Conclusion

This project proves our capability to work together using GitHub, perform meaningful data analysis, and answer real-world business questions using Python and statistics.

Screenshots – Proof of Merge

The following screenshot shows the final state of the repository with all files successfully merged into the main branch. This confirms that all project contributions have been integrated into the main branch, as required.

The screenshot displays the GitHub interface for the repository 'Online-Car-Sale-Analysis'. At the top, it shows the repository name, a 'Public' badge, and interaction buttons for 'Watch' (1), 'Fork' (1), and 'Star' (0). Below this, the repository is viewed on the 'main' branch, showing 7 branches and 0 tags. A search bar and 'Add file' button are visible. The main content area shows a list of commits, with the most recent being a merge pull request #5 from mceballos92/SecPartHQ. The commit history includes files like .ipynb_checkpoints, data, work in Progress_AAlvarez, DS311_Group_Project_Option_4.pdf, Dealer with the most listing and cheapest car ..., Goup project clean up online cars.ipynb, LICENSE, Most popular car by listing by rating by engine..., Online-Cars-Sale-Marketplace-Data-Set-Descri..., README.md, What's the average price for cars with low mile..., and desktop.ini. The right sidebar shows the 'About' section with a description of the repository, 'Readme', 'GPL-2.0 license', 'Activity', '0 stars', '1 watching', '1 fork', and 'Report repository'. Below this are sections for 'Releases' (no releases published), 'Packages' (no packages published), 'Contributors' (2 contributors: alexxx2339 and mceballos92), and 'Languages' (Jupyter Notebook 100.0%). The bottom section shows the 'README' file, which contains the repository's description: 'This repo goes over data on car sale listings from a popular online auto marketplace which identifies interesting/useful information.'

Online-Car-Sale-Analysis

This repo goes over data on car sale listings from a popular online auto marketplace which identifies interesting/useful information.

Visual Analysis and Data Insights

Pull Requests Merged Overview – Confirmation of All Merges

Overview			
5 Active pull requests		0 Active issues	
<div>5</div> Merged pull requests	<div>0</div> Open pull requests	<div>0</div> Closed issues	<div>0</div> New issues

There hasn't been any commit activity on mceballos92/Online-Car-Sale-Analysis in the last week.
Want to help out?

Fork this repository

5 Pull requests merged by 1 person

- Sec part hq

#5 merged 15 minutes ago
- Sec part2

#4 merged 1 hour ago
- High q1

#3 merged 1 hour ago
- Clean data

#2 merged 1 hour ago
- Initial clean up of data and added to README file and answers to question number1.

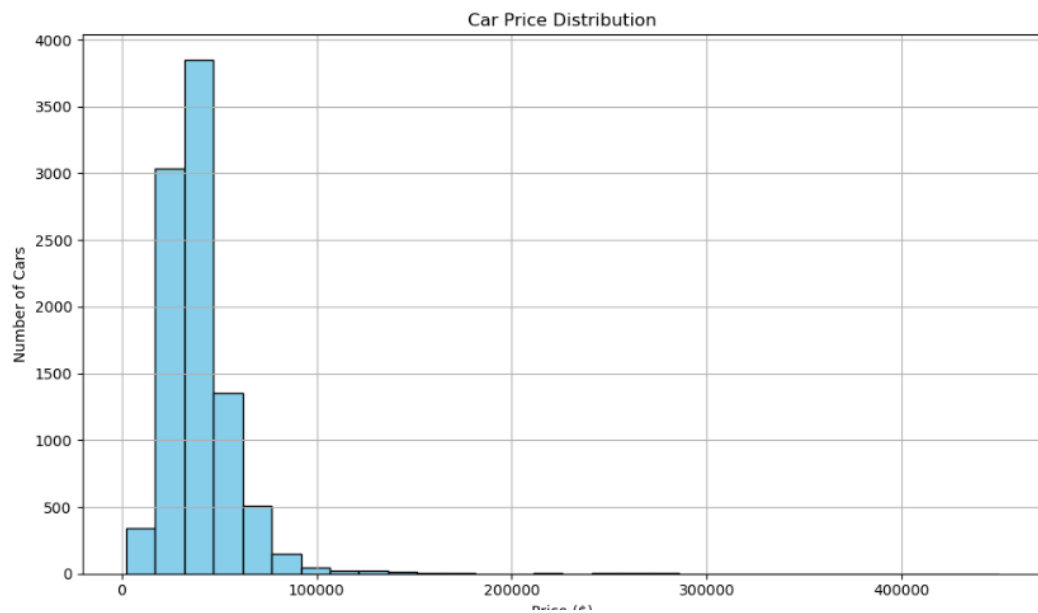
#1 merged 1 hour ago

Car Price Distribution – Histogram of Used Car Prices

```
In [39]: import matplotlib.pyplot as plt

# Make a histogram of car prices
plt.figure(figsize=(10,6))
plt.hist(df['price'].dropna(), bins=30, color='skyblue', edgecolor='black')

plt.title('Car Price Distribution')
plt.xlabel('Price ($)')
plt.ylabel('Number of Cars')
plt.grid(True)
plt.tight_layout()
plt.show()
```



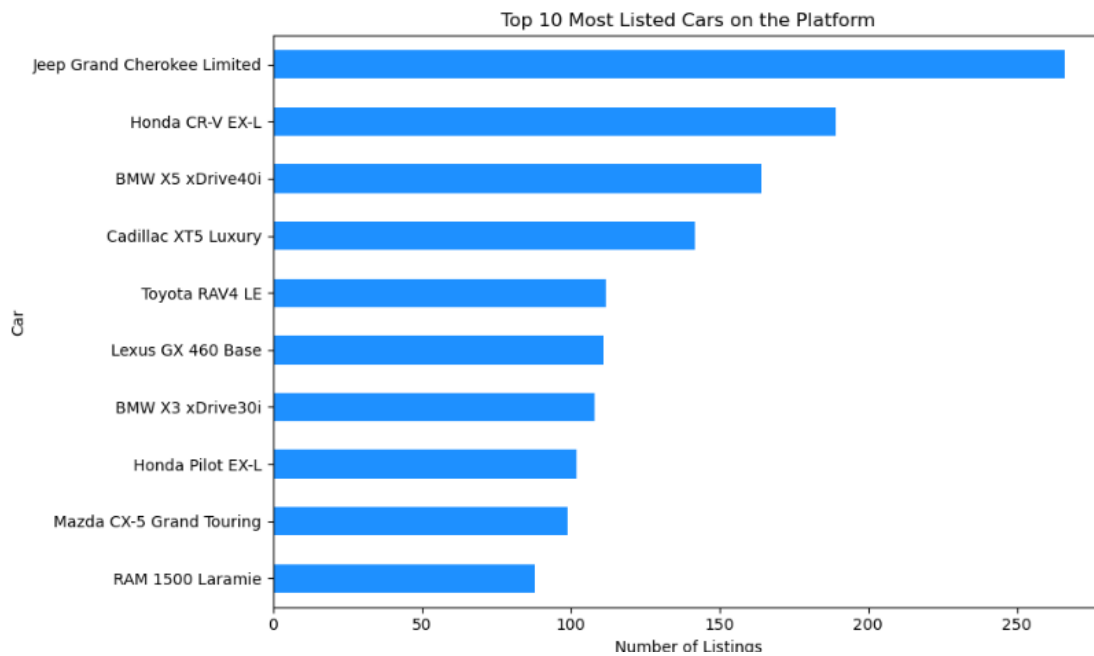
Top 10 Most Listed Cars on the Platform

```
# Create the plot
plt.figure(figsize=(10,6))
popular_listings.plot(kind='barh', color='dodgerblue')

# Titles and labels
plt.title('Top 10 Most Listed Cars on the Platform')
plt.xlabel('Number of Listings')
plt.ylabel('Car')

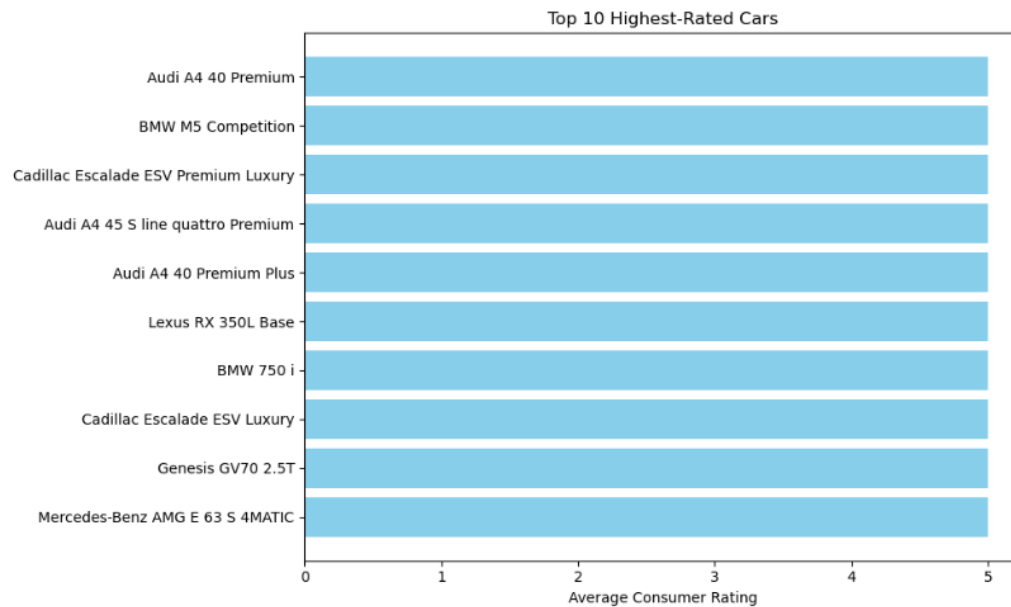
# Invert y-axis so highest is at top
plt.gca().invert_yaxis()

plt.tight_layout()
plt.show()
```



Top 10 Highest-Rated Cars by Average Consumer Rating

```
# Plot
plt.figure(figsize=(10,6))
plt.barh(top_rated['Make'] + ' ' + top_rated['Model'], top_rated['ConsumerRating'], color='skyblue')
plt.xlabel('Average Consumer Rating')
plt.title('Top 10 Highest-Rated Cars')
plt.gca().invert_yaxis() # highest rating on top
plt.tight_layout()
plt.show()
```



Popular by rating

Top 10 Most Common Engine Types in Listings

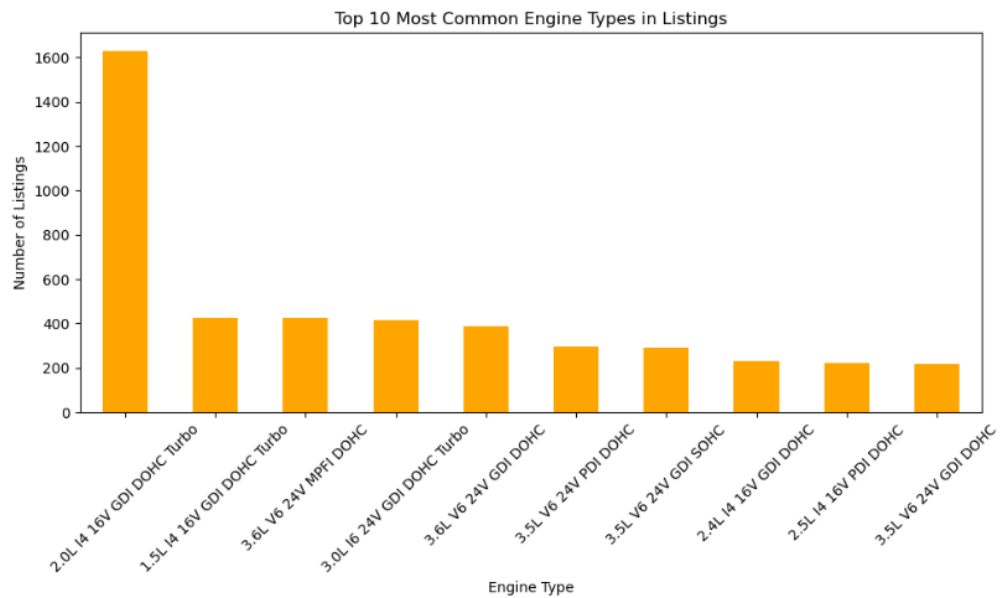
Popular by Engine size

```
In [129... engine_popularity = df['Engine'].value_counts().head(10)

# Display the counts
print(engine_popularity)
```

```
Engine
2.0L I4 16V GDI DOHC Turbo    1629
1.5L I4 16V GDI DOHC Turbo     425
3.6L V6 24V MPFI DOHC         423
3.0L I6 24V GDI DOHC Turbo     417
3.6L V6 24V GDI DOHC          390
3.5L V6 24V PDI DOHC          297
3.5L V6 24V GDI SOHC          290
2.4L I4 16V GDI DOHC          230
2.5L I4 16V PDI DOHC          221
3.5L V6 24V GDI DOHC          216
Name: count, dtype: int64
```

```
In [131... plt.figure(figsize=(10,6))
engine_popularity.plot(kind='bar', color='orange')
plt.title('Top 10 Most Common Engine Types in Listings')
plt.xlabel('Engine Type')
plt.ylabel('Number of Listings')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



Mileage and Price Summary for Cars Under 100,000 Miles

```
In [83]: # Filter for cars with mileage under 100,000
low_mileage_cars = df_clean[df_clean['mileage'] < 100000]

# Display first few rows
print(low_mileage_cars[['make', 'model', 'year', 'mileage', 'price']].head())
```

	make	model	year	mileage	price
0	Toyota	Sienna SE	2019	29403	\$39,998
1	Ford	F-150 Lariat	2018	32929	\$49,985
2	RAM	1500 Laramie	2017	23173	\$41,860
3	Honda	Accord Sport SE	2021	10598	\$28,500
4	Lexus	RX 350	2020	28137	\$49,000

```
In [85]: print(df_clean['mileage'].describe())
print(df_clean['mileage'].isnull().sum())
```

```
count      9379.000000
mean       37463.023350
std        24970.342569
min         121.000000
25%        18666.500000
50%        32907.000000
75%        47698.000000
max        234114.000000
Name: mileage, dtype: float64
0
```