THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Exploring and Prediction of Used-Car Market

EMSE 6574

Instructor: Dr. Maksim Tsvetovat

Group Member:

Rui Zhang

Ran Wei

Bite Xiong

Fangzhou Liu

Contents

- Introduction and Data Set (fangzhou)
- Exploratory Data Analysis (Ran Wei)
- Data Preprocessing (Rui Zhang)
- Prediction Model (Bite Xiong)
- Conclusion

Introduction and Data Set

- Create a method to determine whether the price of a used car is appropriate

- Cleaning Data sets for 2020 and 2021
  - Remove unnecessary features
  - Determine the target price and years(from $320 to $105,000)
  - Remove outliers (3-sigma rule)

Cleaned 2020 data set

```
cars_2020.describe()
```

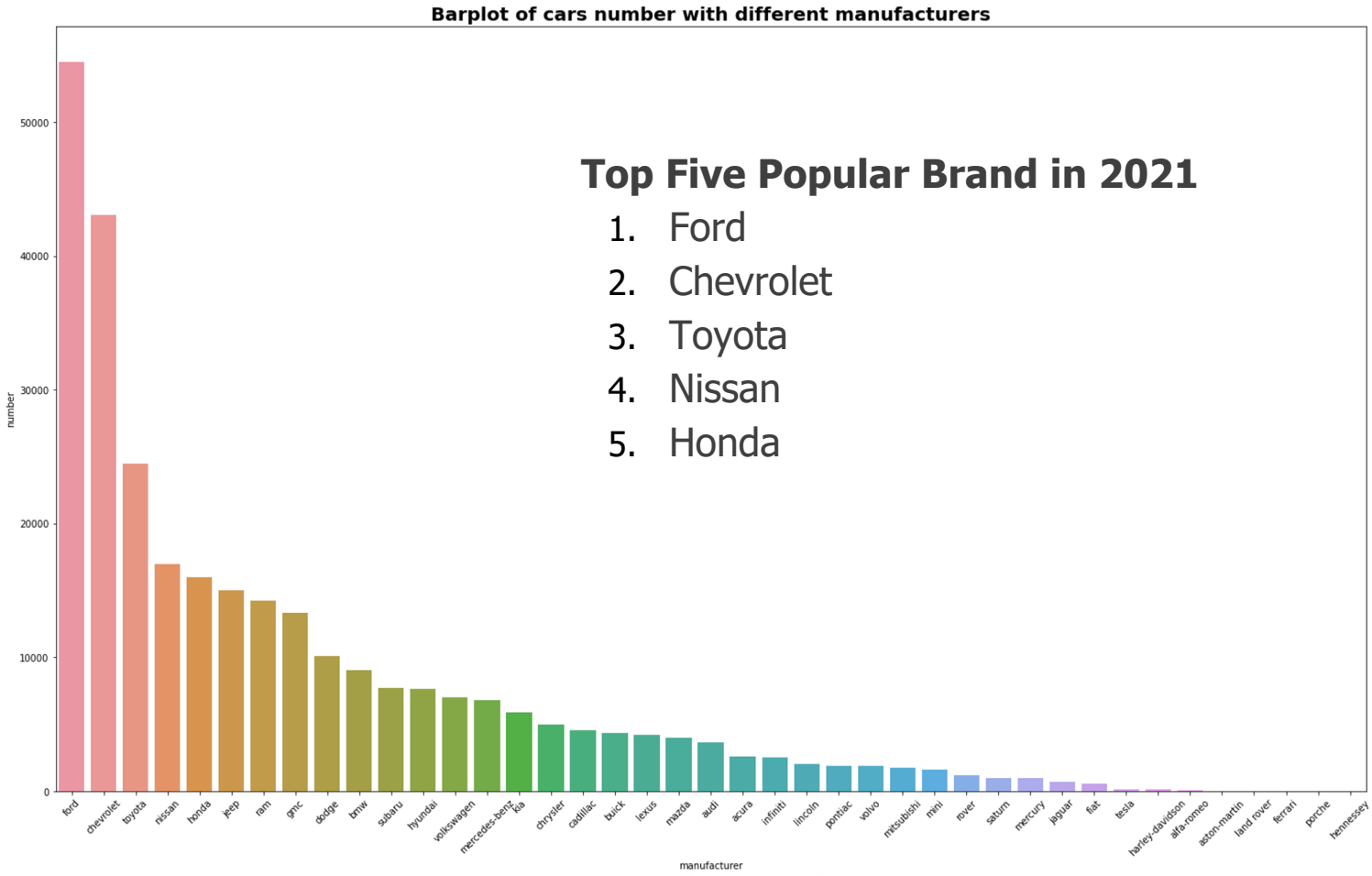
	price	year	odometer
count	305283.000000	305283.000000	305283.000000
mean	14018.423466	2010.893820	100559.474294
std	10610.744487	5.622829	61812.419148
min	311.000000	1990.000000	0.000000
25%	5995.000000	2007.000000	49500.500000
50%	11500.000000	2012.000000	96658.000000
75%	18995.000000	2015.000000	141007.500000
max	83000.000000	2020.000000	399961.000000

Cleaned 2021 data set

```
cars_2021.describe()
```

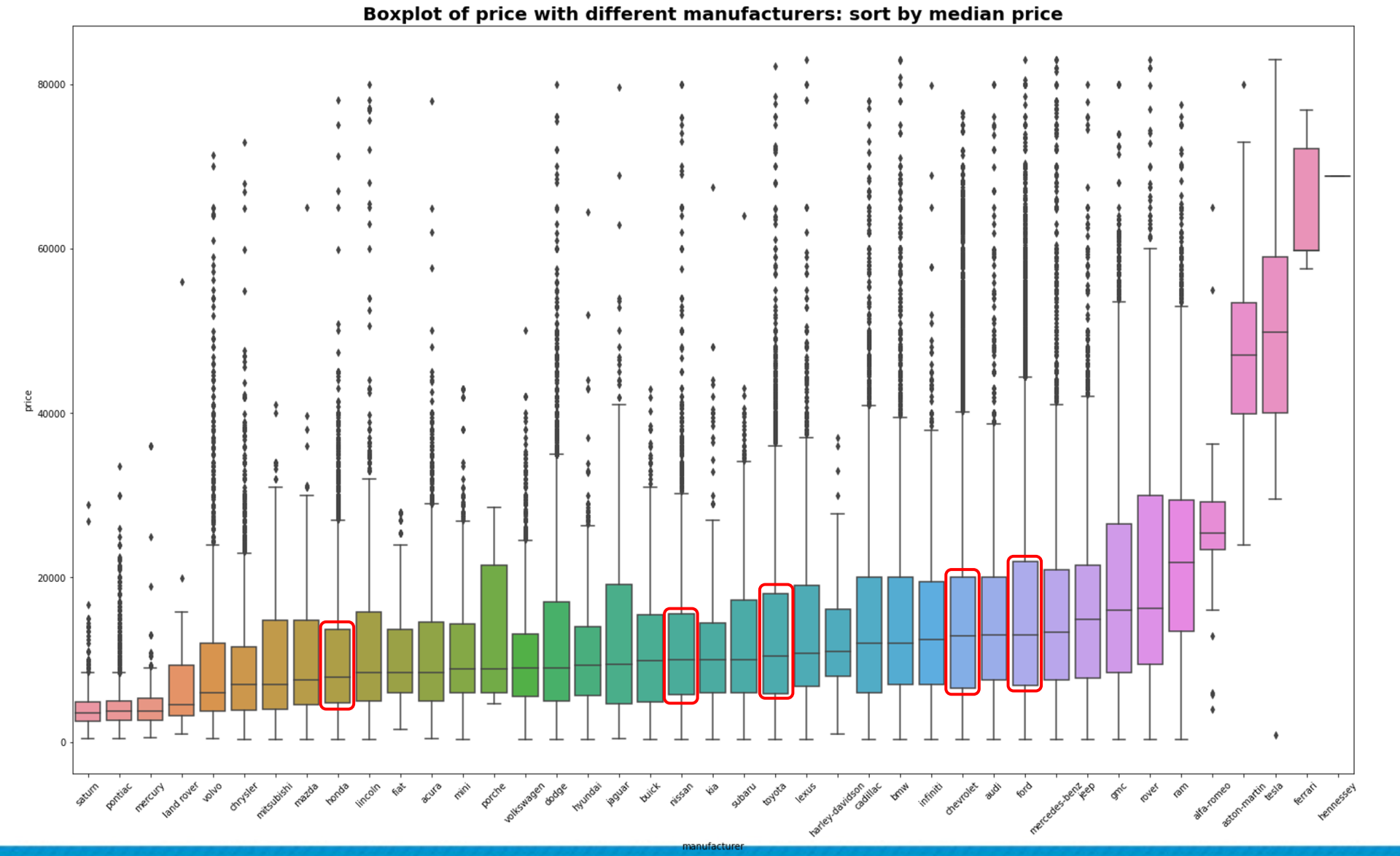
	price	year	odometer
count	259962.000000	259962.000000	259962.000000
mean	17806.449650	2011.745986	97996.685058
std	14132.519111	6.009529	64559.348267
min	320.000000	1991.000000	0.000000
25%	6950.000000	2008.000000	42548.000000
50%	13996.000000	2013.000000	93988.000000
75%	25900.000000	2017.000000	141199.250000
max	103900.000000	2021.000000	399999.000000

Exploratory Data Analysis

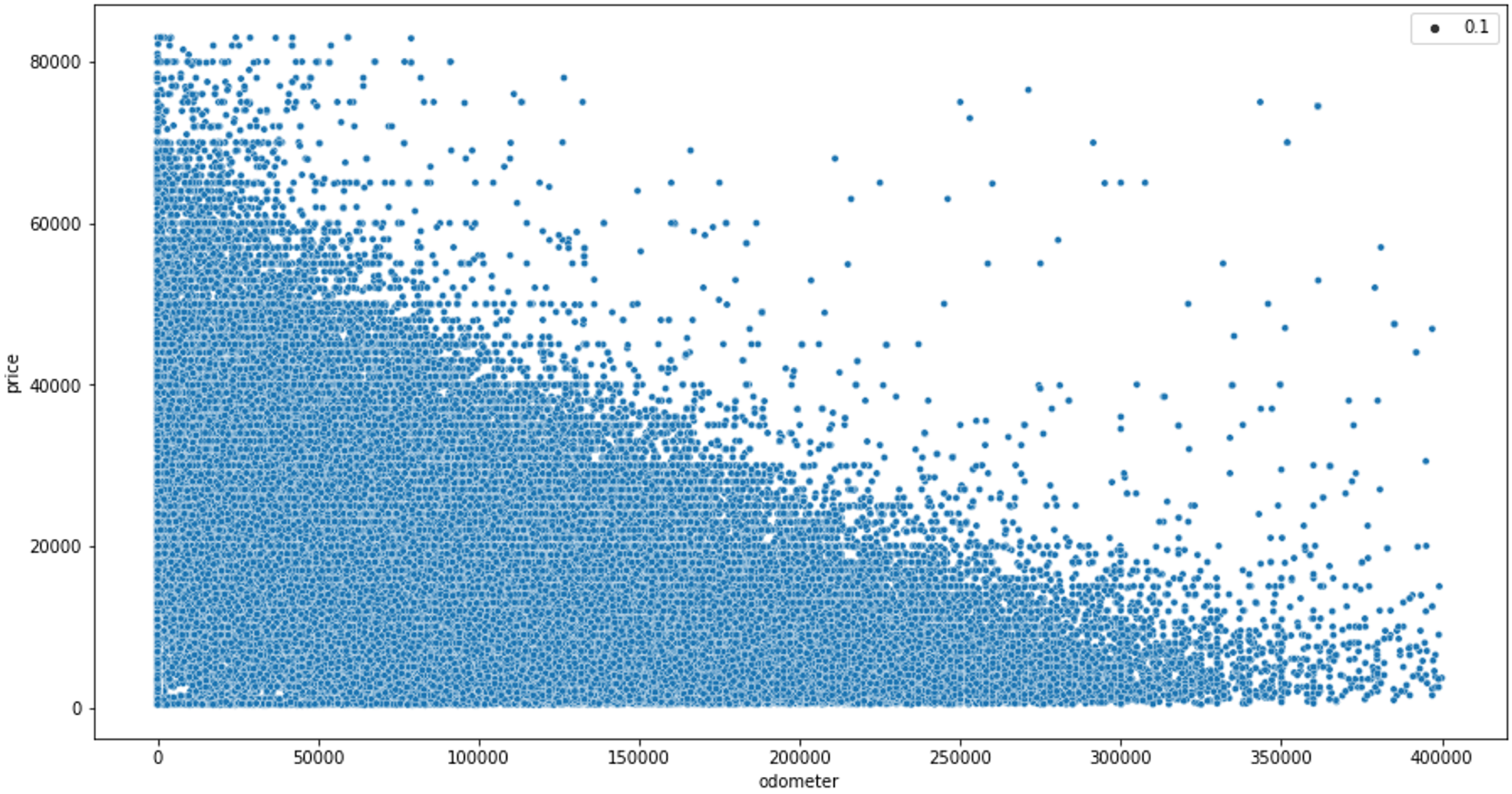


Top Five Popular Brand in 2021

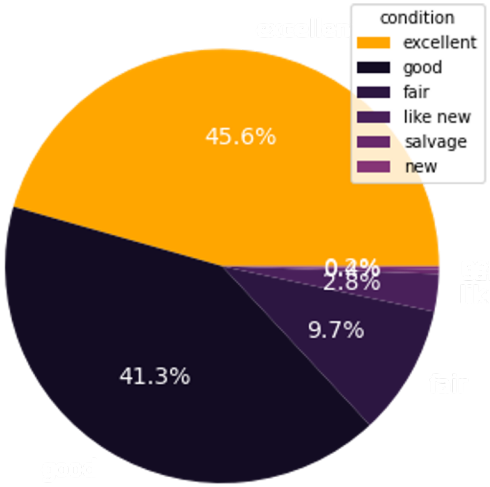
1. Ford
2. Chevrolet
3. Toyota
4. Nissan
5. Honda

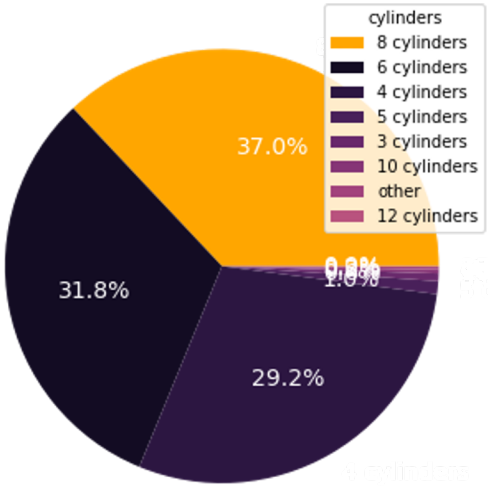


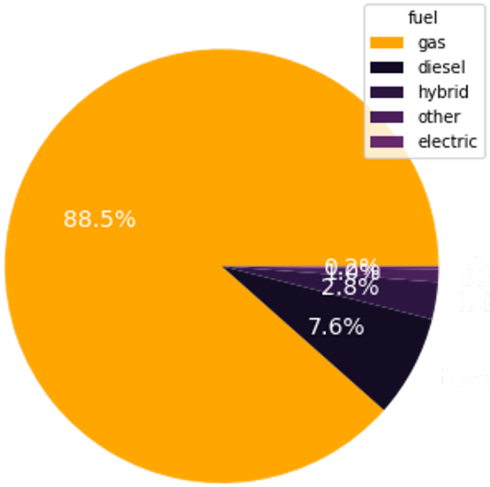
Scatter plot for price and odometer

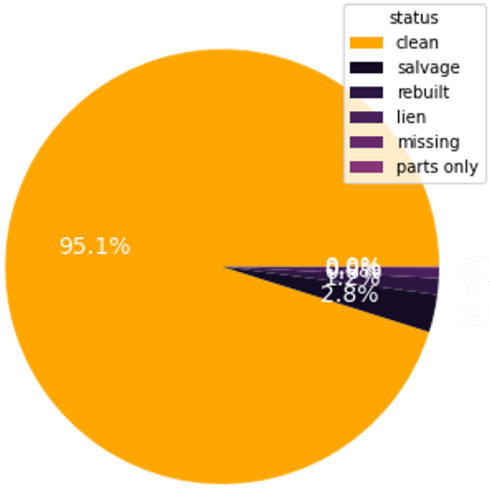


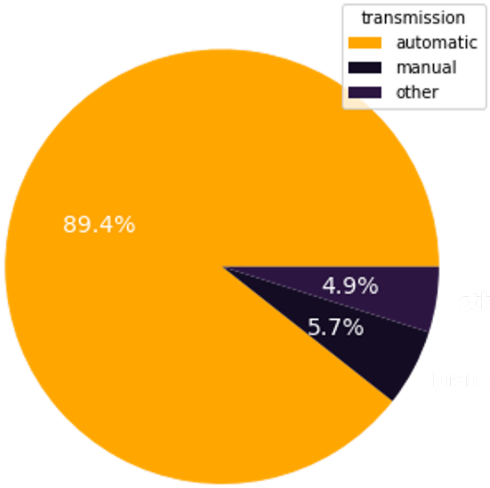
The proportion of values in each category features

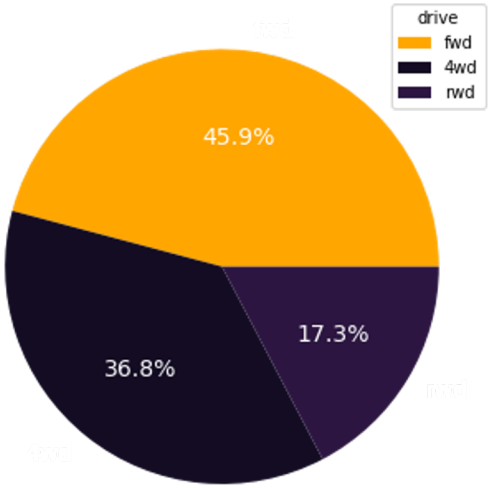




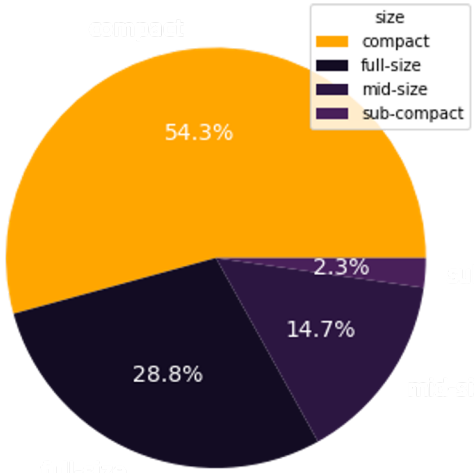


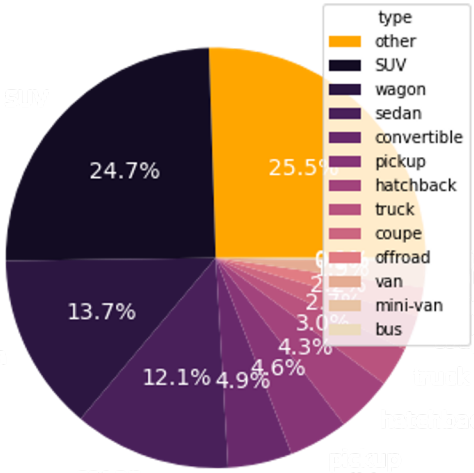


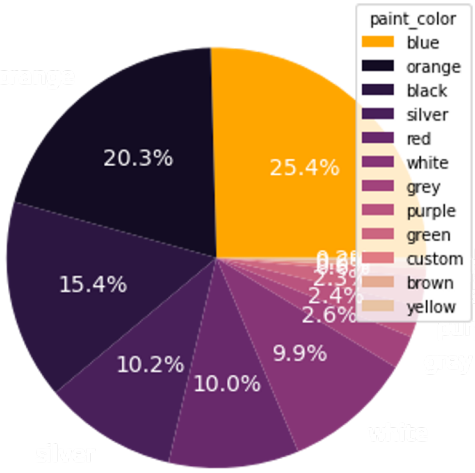




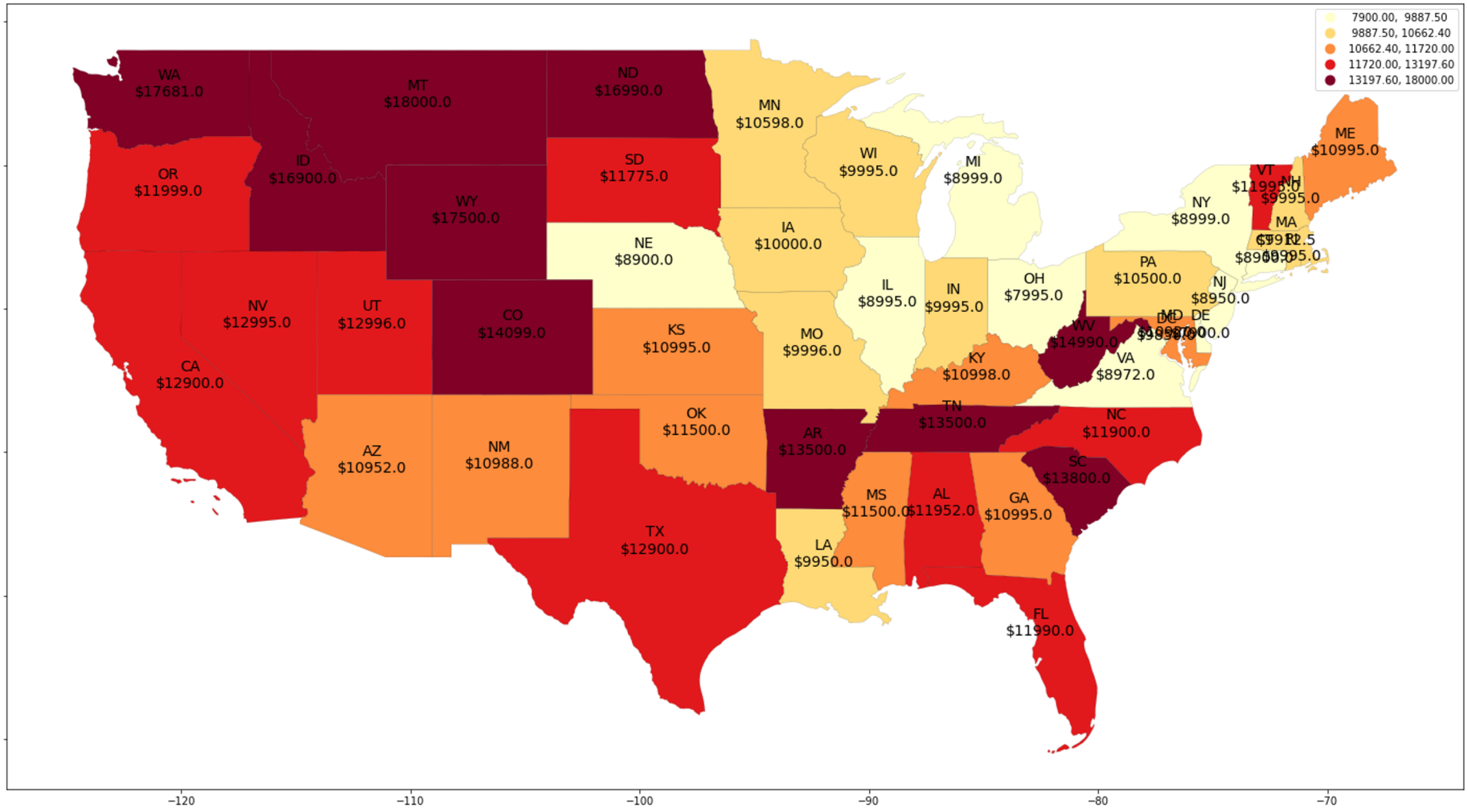
The proportion of values in each category features



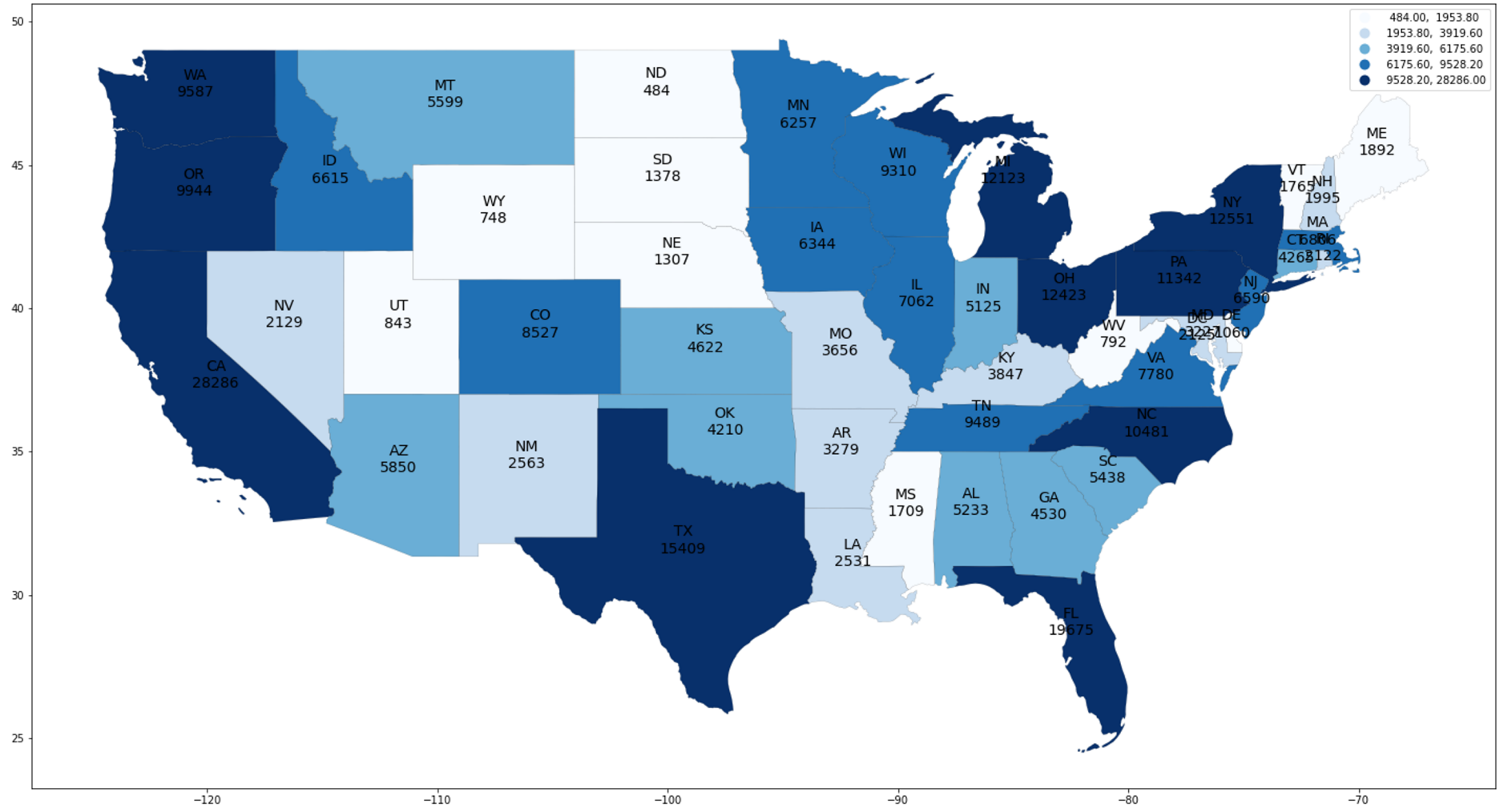


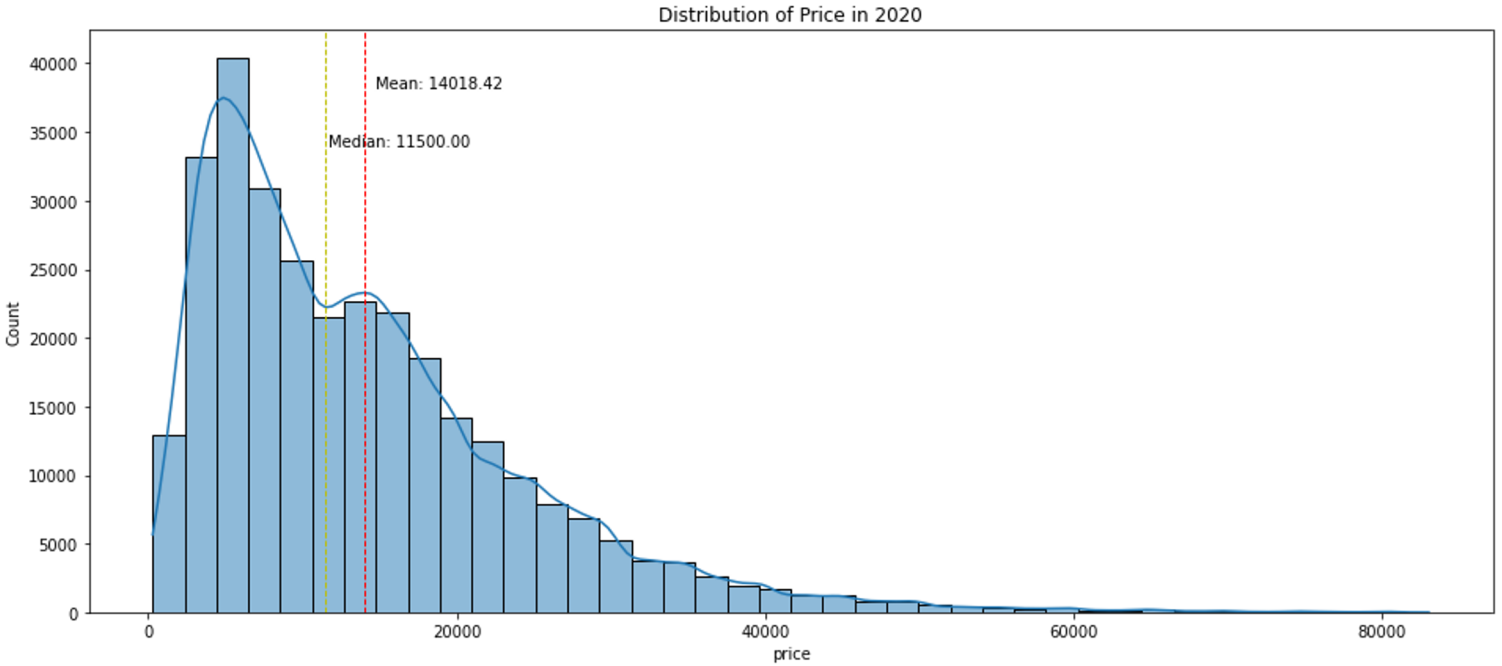


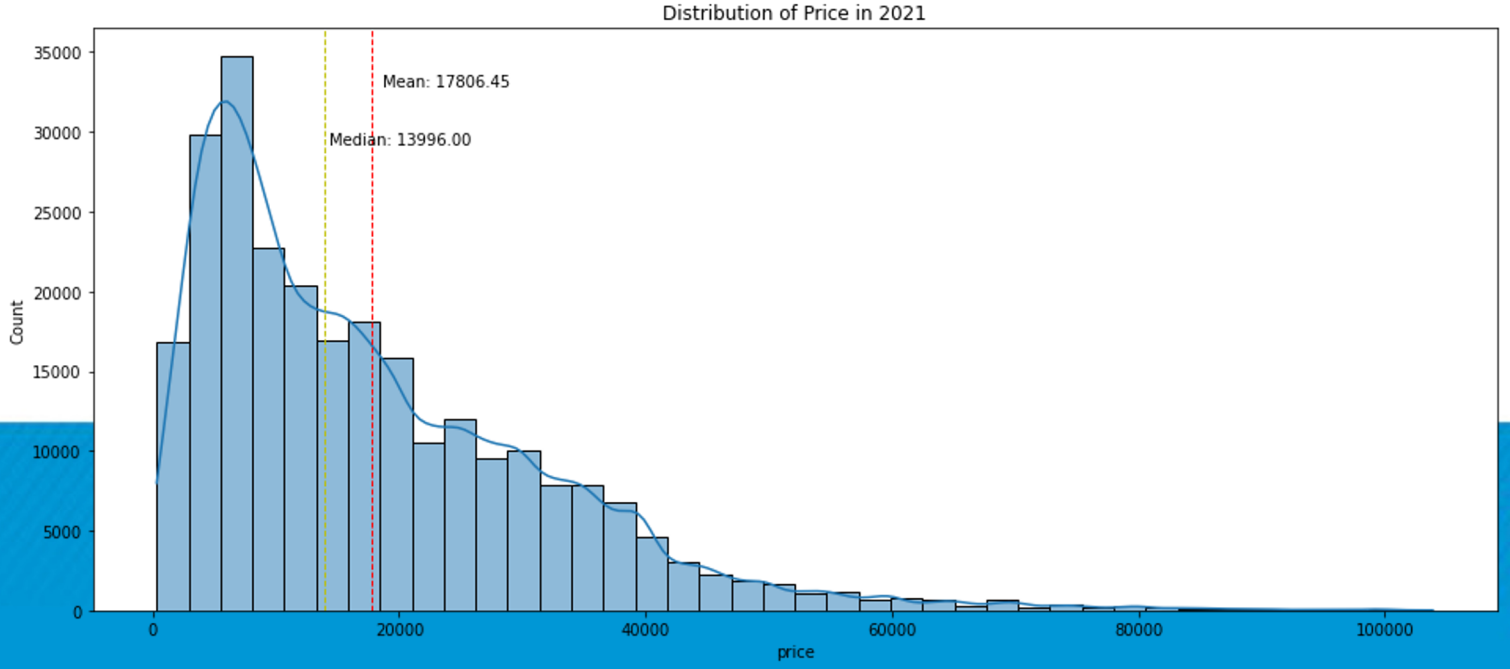
Median Price of Used Cars in Each State



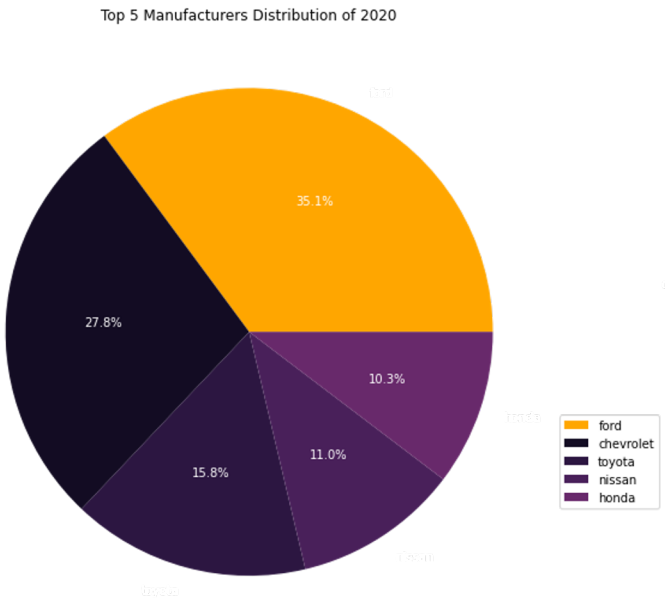
Numbers of Used Cars in Each State

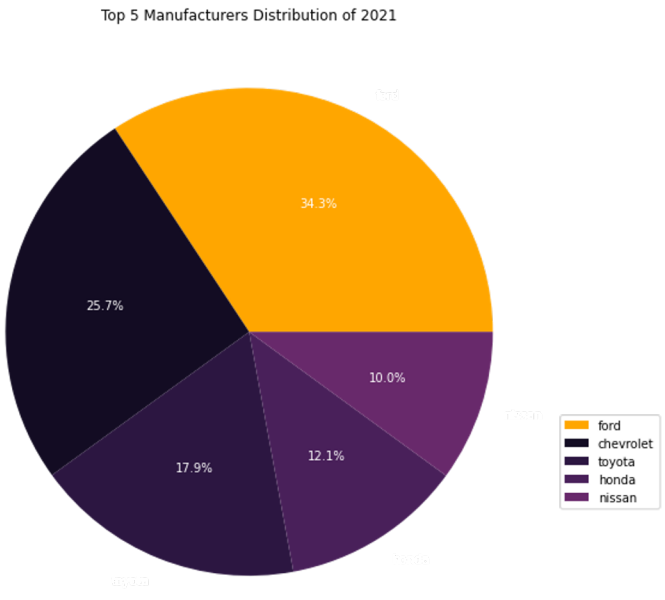






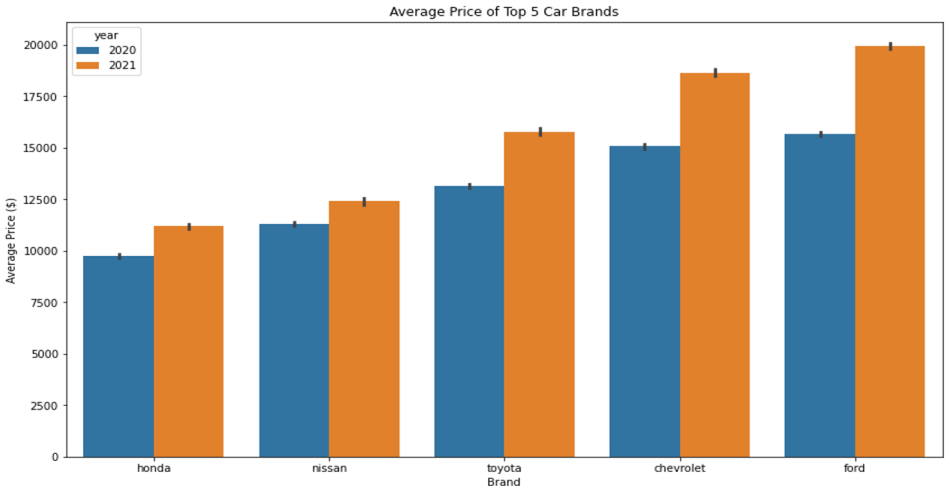
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC





Used-Car Market

2021 V.S. 2020



THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Data Preprocessing

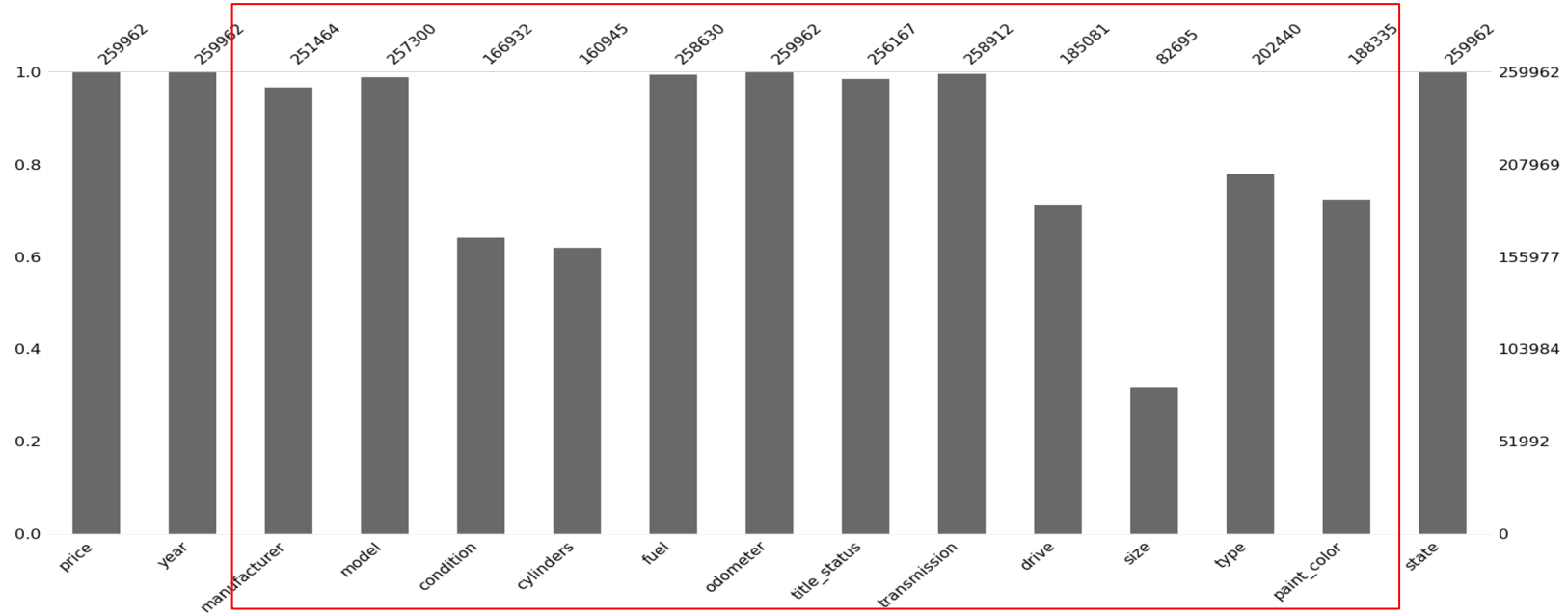
- Missing Value
- Categorical Features Encoding
- Normalize Numerical Features

Data Preprocessing

➢ Missing Value

Problem: Most of the categorical data have missing values.

Solution: Fill missing values with”UNKNOWN”.

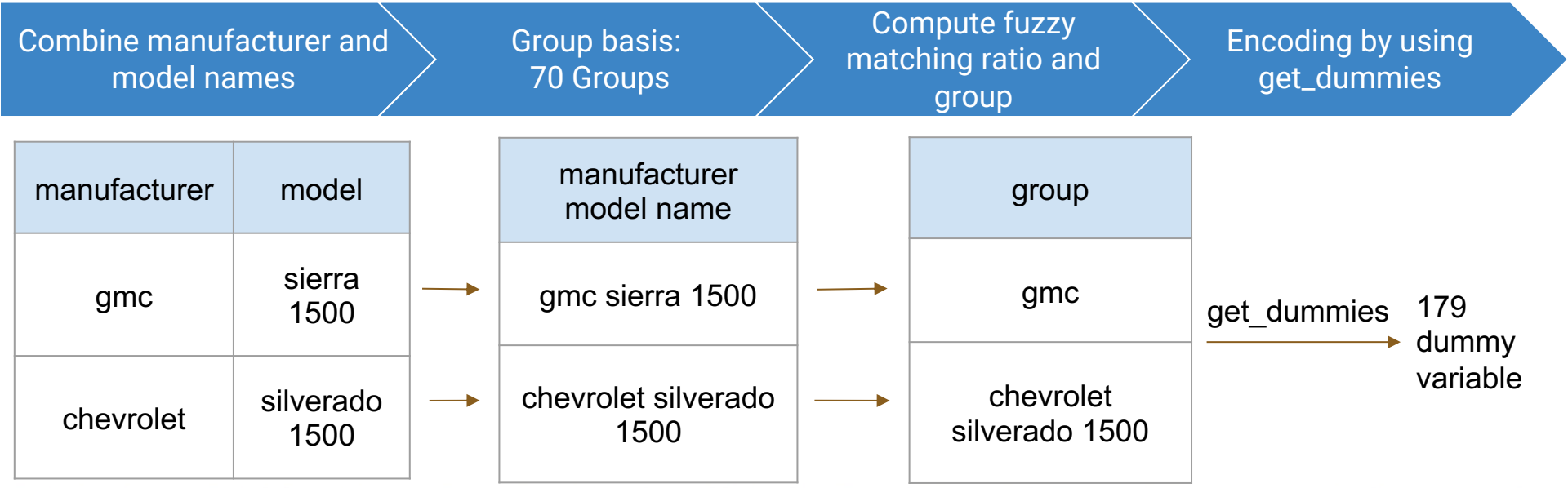


Data Preprocessing

➢ Categorical Features Encoding

Problem: More than 27,000 car model names lead to too many features after One-Hot Encoding.

Solution: Group car model names by fuzzy matching.



Data Preprocessing

➢ Normalize Numerical Features

Problem: Some of numerical features have different ranges, may affect model results.

Solution: Normalize numerical features by using MinMaxScaler.

Prediction Model

- Train Test Split
- Fit Model
- Good/Bad Deal & Find a Best Car

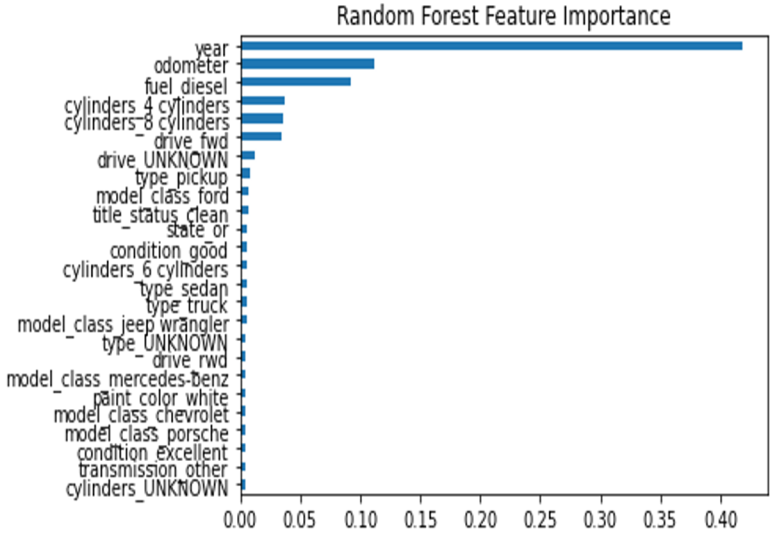
Prediction Model

➢ Train Test Split

Train Data	174174 (67%)
Test Data	85788 (33%)

➢ Fit Model

	RMSE_train	RMSE_test	R2_train	R2_test
Linear Regression	8272.25	8338.43	0.6134	0.6082
Decision Tree	364.99	7533.07	0.9993	0.7164
Random Forest Regression	2138.58	5756.26	0.9771	0.8344
XGBoost Regression	7448.27	7580.64	0.722	0.7128



Prediction Model

➢ Good/Bad Deal & Find a Best Car

- Determine whether it is good deal or bad deal

	model_class_jeep wrangler	predict price	price	Good/Bad Deal
245634	1	35522.05	36683	Bad Deal
215405	1	31712.00	31590	Good Deal
22052	1	40517.00	39990	Good Deal
44404	1	10061.43	10800	Bad Deal
185489	1	20775.28	9500	Good Deal

- Find the best car in Lexus
- Find the best jeep wrangler

```
Best_Car('model_class_lexus')
```

	price	year	manufacturer	model	condition	cylinders	fuel	odometer	title_status	transmission	drive	size	type	paint_color	state
66877	1200	2013	lexus	es 350	excellent	6 cylinders	gas	92699.0	clean	automatic	fwd	UNKNOWN	sedan	UNKNOWN	fl

```
Best_Car('model_class_jeep wrangler')
```

	price	year	manufacturer	model	condition	cylinders	fuel	odometer	title_status	transmission	drive	size	type	paint_color	state
5653	2000	2015	jeep	wrangler jku	good	UNKNOWN	gas	1.0	clean	other	UNKNOWN	UNKNOWN	UNKNOWN	UNKNOWN	az

CONCLUSION

Final Paper:

https://docs.google.com/document/d/1F4m_1mJYTmVqU45oO-i5lj7URbOjNtJG1g_fH58TMzs/edit

Github Link:
https://github.com/Melody1745/Used_Car_Market_Analysis.git

Colab Link:

https://drive.google.com/drive/folders/1itFGWLRWcPtWGKpcqWCopesh87L0HBUw?usp=sharing

Thank You
&
Questions ?

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC