



Craigslist Project

Analysis of Used Car Sales

Team 14: Siyi Chen, Wensi Ouyang, Ray Chang



Table of Contents

01

About
Craigslist

02

Original Data

03

Data Cleaning

04

Examine the
Clean Data

05

Research
Question

06

Trends and
Insights



Craigslist is an American classified advertisements website with sections devoted to jobs, housing, for sale, items wanted, services, community service, gigs, résumés, and discussion forums.

Craigslist is an American classified advertisements website with sections devoted to jobs, housing, for sale, items wanted, services, community service, gigs, résumés, and discussion forums.

craigslist

[create a posting](#)

[my account](#)

search craigslist

event calendar						
M	T	W	T	F	S	S
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

[craigslist app](#) new

[help, faq, abuse, legal](#)

[avoid scams & fraud](#)

[personal safety tips](#)

[terms of use](#)

[privacy policy](#) new

[system status](#)

[about craigslist](#)

[craigslist is hiring in sf](#)

[craigslist open source](#)

[craigslist blog](#)

[best-of-craigslist](#)

"[craigslist joe](#)"

bloomington, IN

community

- [activities](#)
- [artists](#)
- [childcare](#)
- [classes](#)
- [events](#)
- [general](#)
- [groups](#)
- [local news](#)
- [lost+found](#)
- [missed connections](#)
- [musicians](#)
- [pets](#)
- [politics](#)
- [rants & raves](#)
- [rideshare](#)
- [volunteers](#)

services

- [automotive](#)
- [beauty](#)
- [cell/mobile](#)
- [computer](#)
- [creative](#)
- [cycle](#)
- [event](#)
- [farm+garden](#)
- [financial](#)
- [health/well](#)
- [household](#)
- [labor/move](#)
- [legal](#)
- [lessons](#)
- [marine](#)
- [pet](#)
- [real estate](#)
- [skilled trade](#)
- [sm biz ads](#)
- [travel/vac](#)
- [write/ed/tran](#)

discussion forums

- [apple](#)
- [arts](#)
- [atheist](#)
- [autos](#)
- [beauty](#)
- [bikes](#)
- [frugal](#)
- [gaming](#)
- [garden](#)
- [haiku](#)
- [help](#)
- [history](#)
- [philos](#)
- [photo](#)
- [politics](#)
- [psych](#)
- [recover](#)
- [religion](#)

housing

- [apts / housing](#)
- [housing swap](#)
- [housing wanted](#)
- [office / commercial](#)
- [parking / storage](#)
- [real estate for sale](#)
- [rooms / shared](#)
- [rooms wanted](#)
- [sublets / temporary](#)
- [vacation rentals](#)

for sale

- [antiques](#)
- [appliances](#)
- [arts+crafts](#)
- [atv/utv/sno](#)
- [auto parts](#)
- [aviation](#)
- [baby+kid](#)
- [barter](#)
- [beauty+hlt](#)
- [bike parts](#)
- [bikes](#)
- [boat parts](#)
- [boats](#)
- [books](#)
- [business](#)
- [cars+trucks](#)
- [cds/dvd/vhs](#)
- [farm+garden](#)
- [free](#)
- [furniture](#)
- [garage sale](#)
- [general](#)
- [heavy equip](#)
- [household](#)
- [jewelry](#)
- [materials](#)
- [motorcycle parts](#)
- [motorcycles](#)
- [music instr](#)
- [photo+video](#)
- [rvs+camp](#)
- [sporting](#)
- [tickets](#)
- [tools](#)

Original Data Set

Number of Rows: 426880

Number of Columns: 26

Number of columns have missing value: 20

Categorical Variables:

id, url, region, region url, manufacturer, model, condition, manufacturer, fuel, cylinders, drive, size, paint_color, title_status, VIN, county, image_url, description, state

Numerical Variables:

Price, year, odometer, lat, long, posting date

Variables that have significant missing values(more than 10k):

drive, size, paint_color, cylinders, condition, VIN, type, county

Data Cleaning

Missing Values(NaN)

paint_color



Fill NaN with unknown

odometer



Fill NaN with unknown

manufacturer



Fill NaN with unknown

model



Fill NaN with unknown

cylinders



Fill NaN with other

fuel



Fill NaN with other

title_status



Fill NaN with other

transmission



Fill NaN with other

drive



Fill NaN with other

condition*



Use different values

of odometer as threshold to
determine conditions

*please refer to the note

Data Cleaning

Unuseful/invalid information

Type

Drop NaN values

Size

Drop entire column (not useful in our future analysis)

*price = 0

Drop rows with price = 0 (invalid posting)

*year

Drop years that exceed 2021

county, url, rigon url, image url

Drop entire columns(not useful in our future analysis)

*rows have 10 or more NA value

Drop entire rows

Data Cleaning

Duplicates

VIN

Since VIN value should be unique for individual cars, we dropped the duplicate values

Adding Column

Age

We use today's date and year of posting to derive the age of car to give us valuable metrix to analysis.

Date Type

posting date

We slice out the posting date and convert it into datetime for future calculation

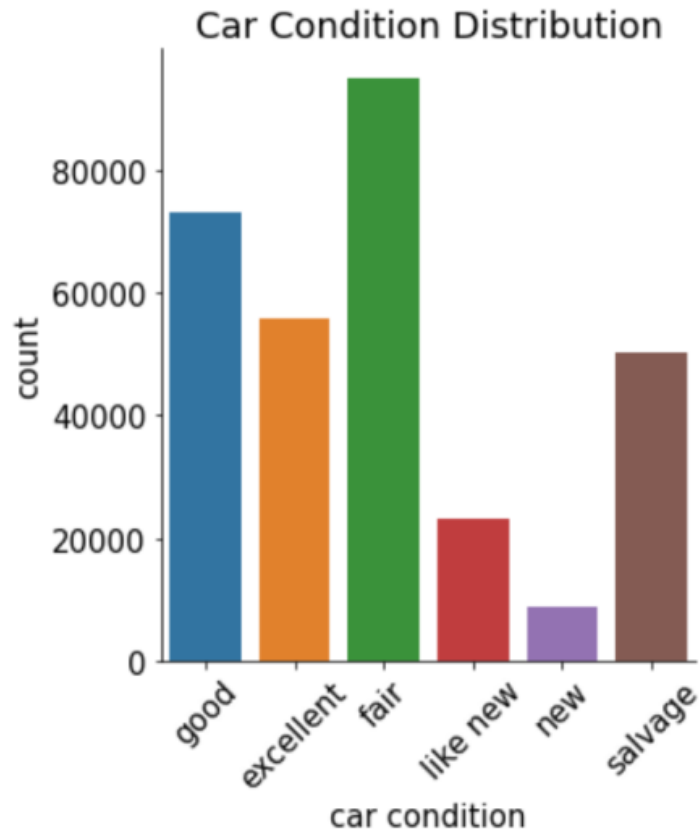
New Data Frame

Number of rows	308004
Number of columns	23

Examine the Clean Data

Conditions of cars & Number of cars in each condition

car condition	
fair	95049
good	73194
excellent	55905
salvage	50176
like new	23114
new	8760



Number of Ads by States/Manufacturer

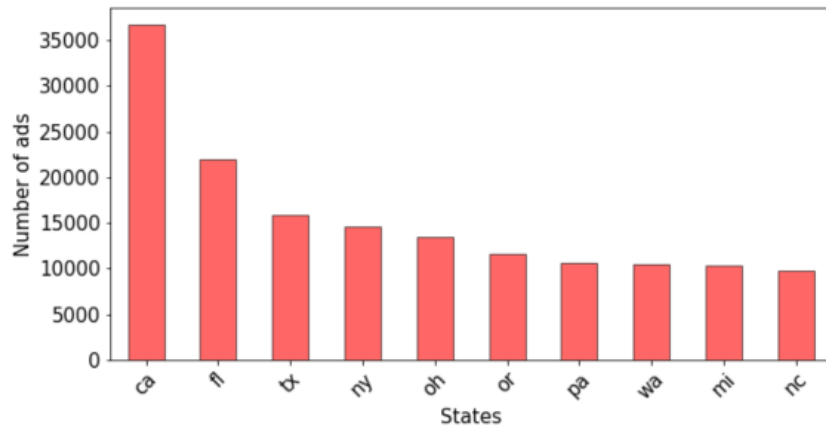
States covered in the dataset:

51 states

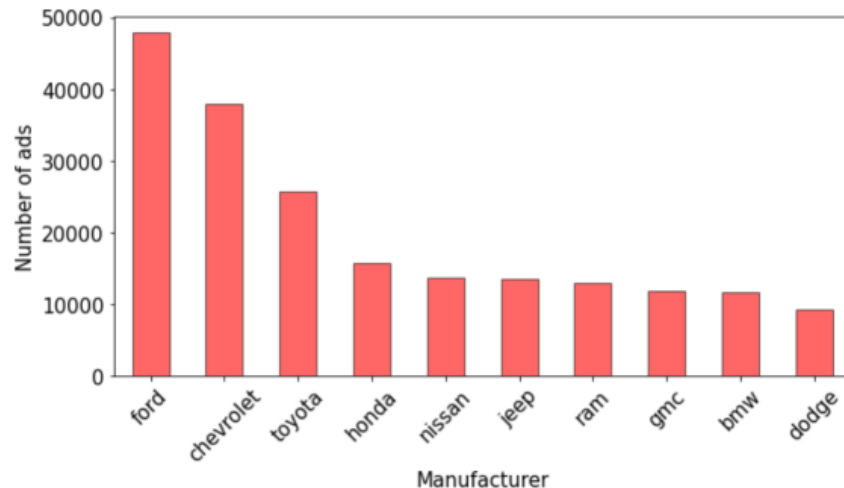
Manufacturers covered in the dataset:

43 manufacturers (with 1 unknown)

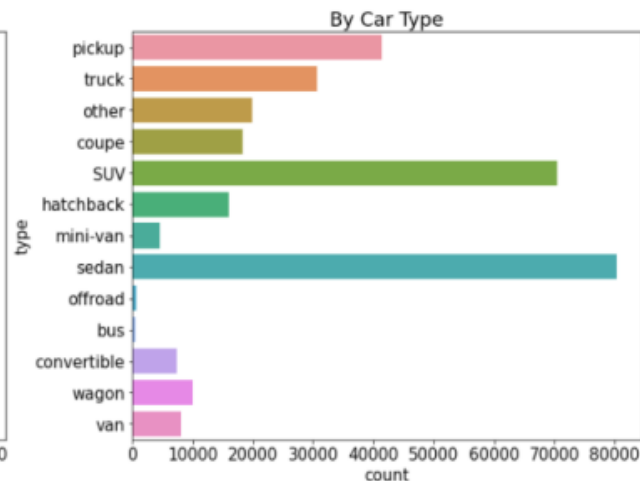
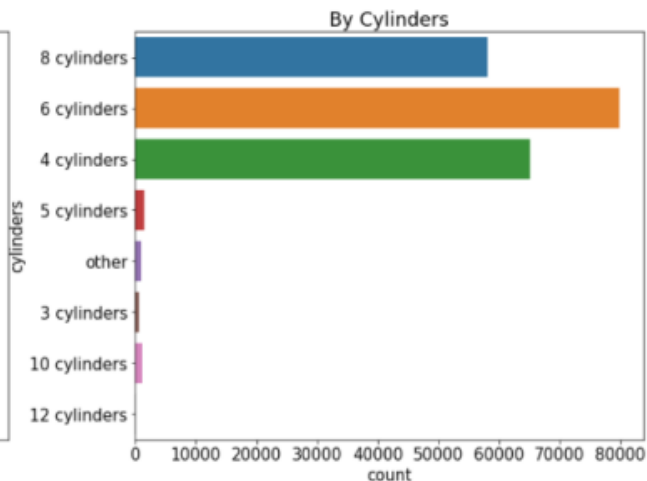
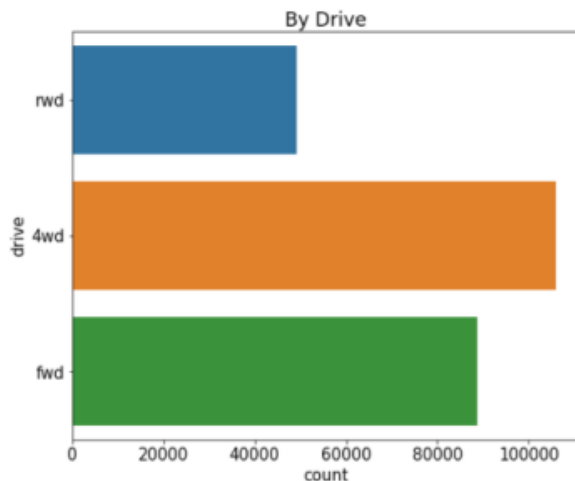
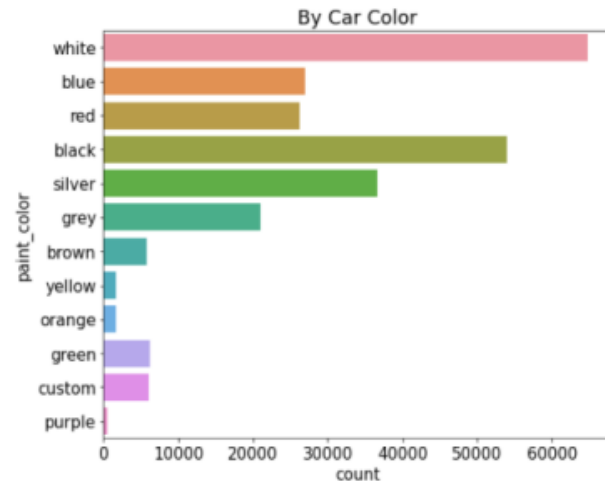
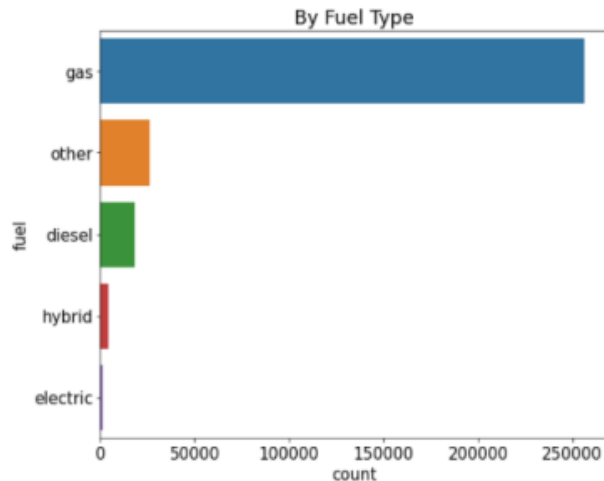
Top 10 Number of Ads by States



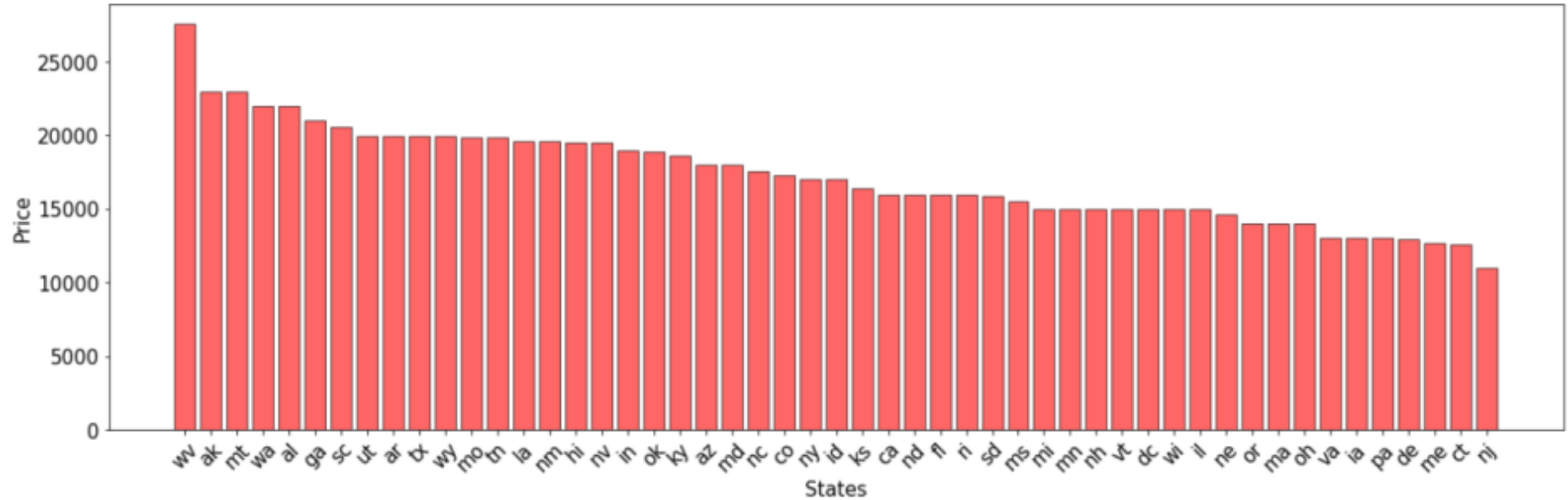
Top 10 Number of Ads by Manufacturer



Compare Number of Cars by different categorical variables

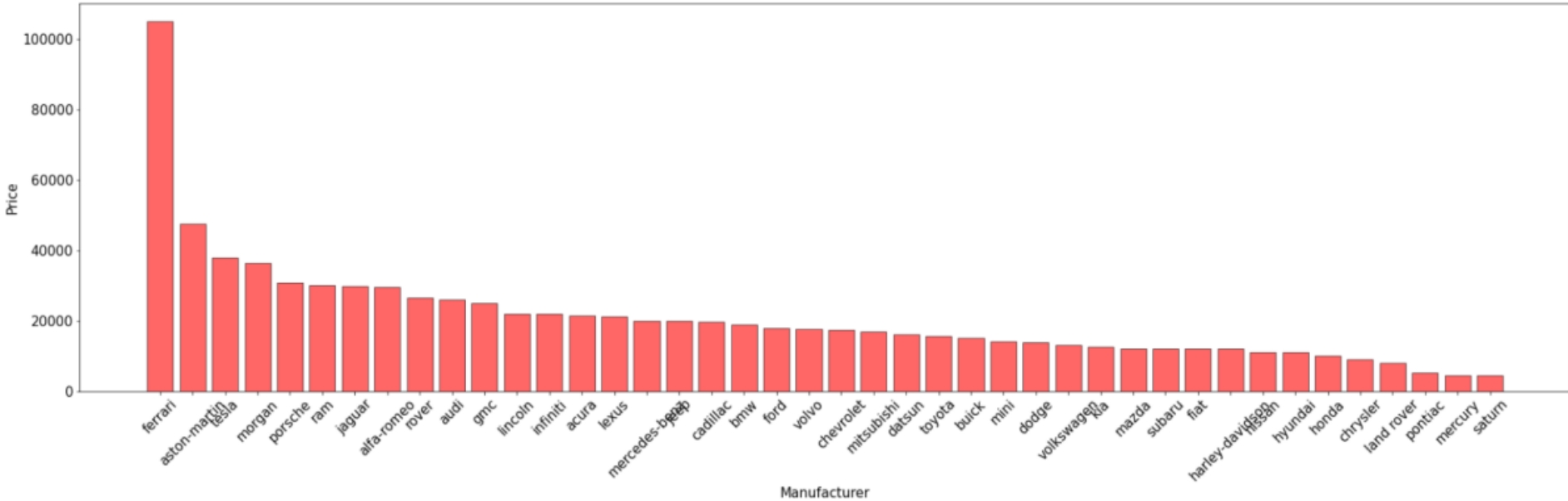


Median Price by State



Based on the analysis, the top three states that has a higher median car price are West Virginia, Alaska, Montana, while New Jersey, Connecticut, and Maine are states that have the lowest median prices for cars.

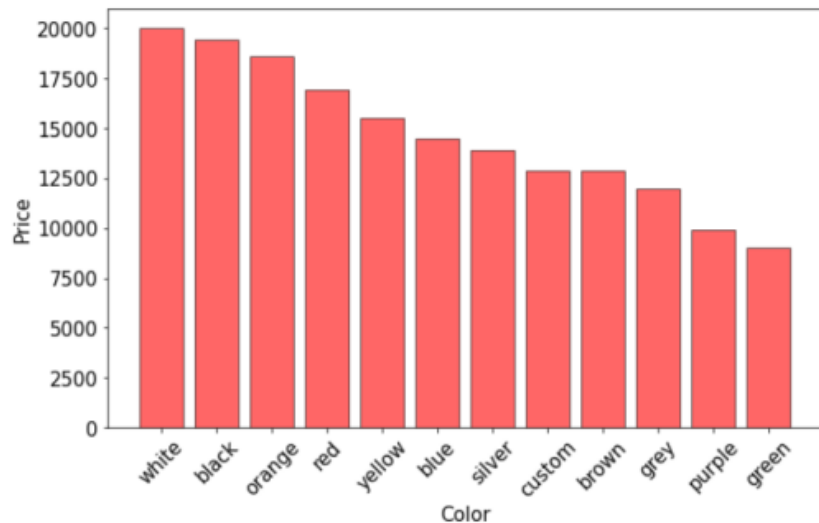
Median Price by Manufacturer



Ferrari, Aston-martin, and Tesla are the most expensive brands. On the other hand, Pontiac, Mercury, and Saturn are manufacturers that sell the least expensive cars among 42 manufacturers.



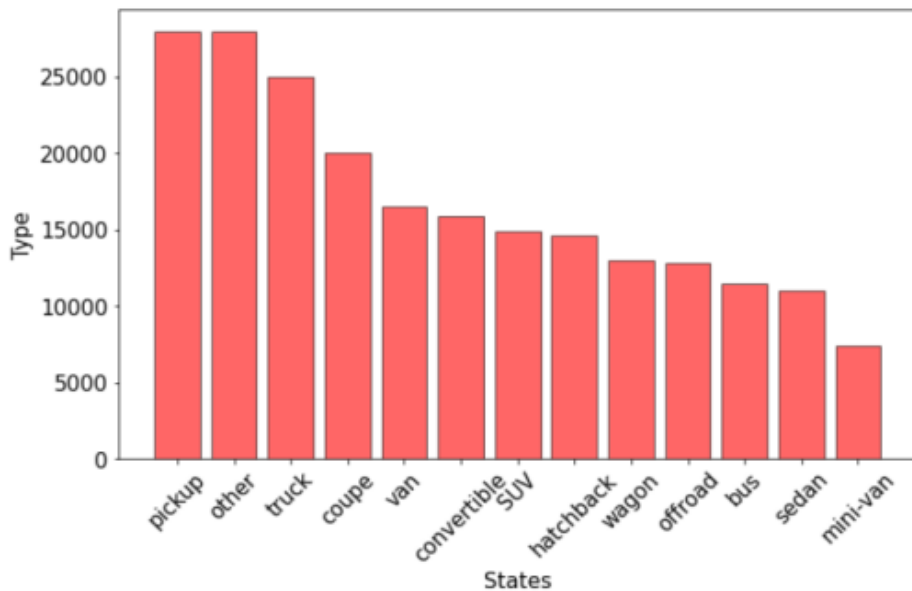
Median Price by Color



White and Black cars are the most expensive ones. If you buy purple or green car, you can expect to have a lower price tag than its white or black model.

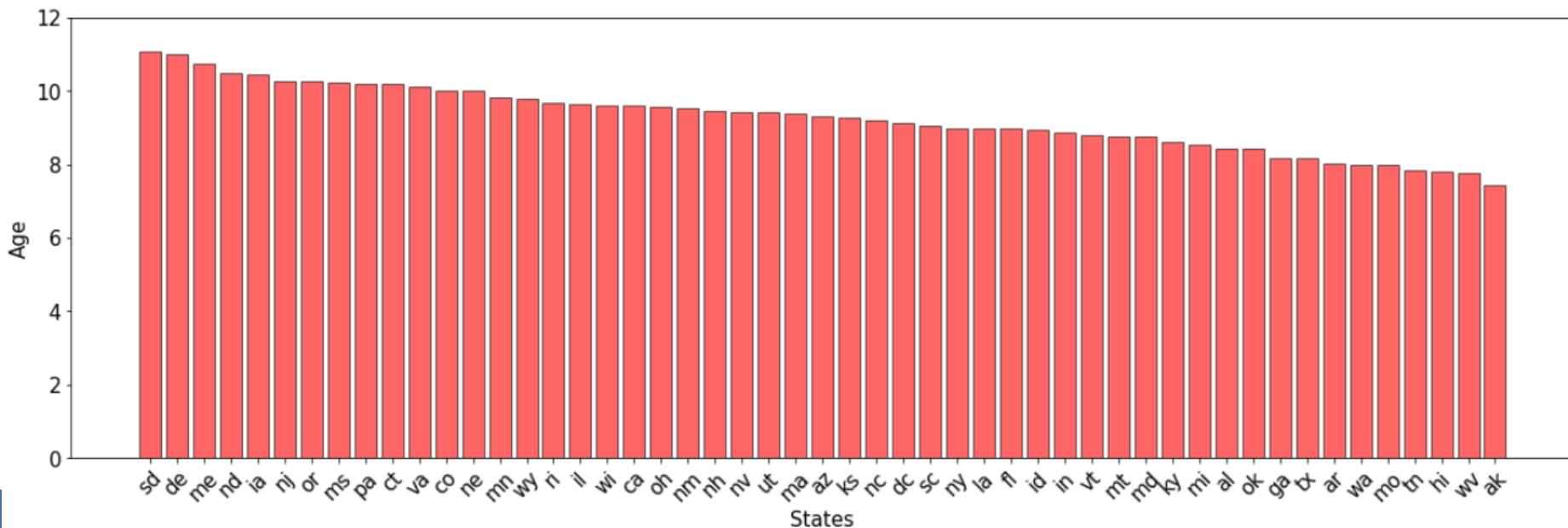
Mini-van is the least expensive car type, and Pickup is the most expensive car type.

Median Price by Type

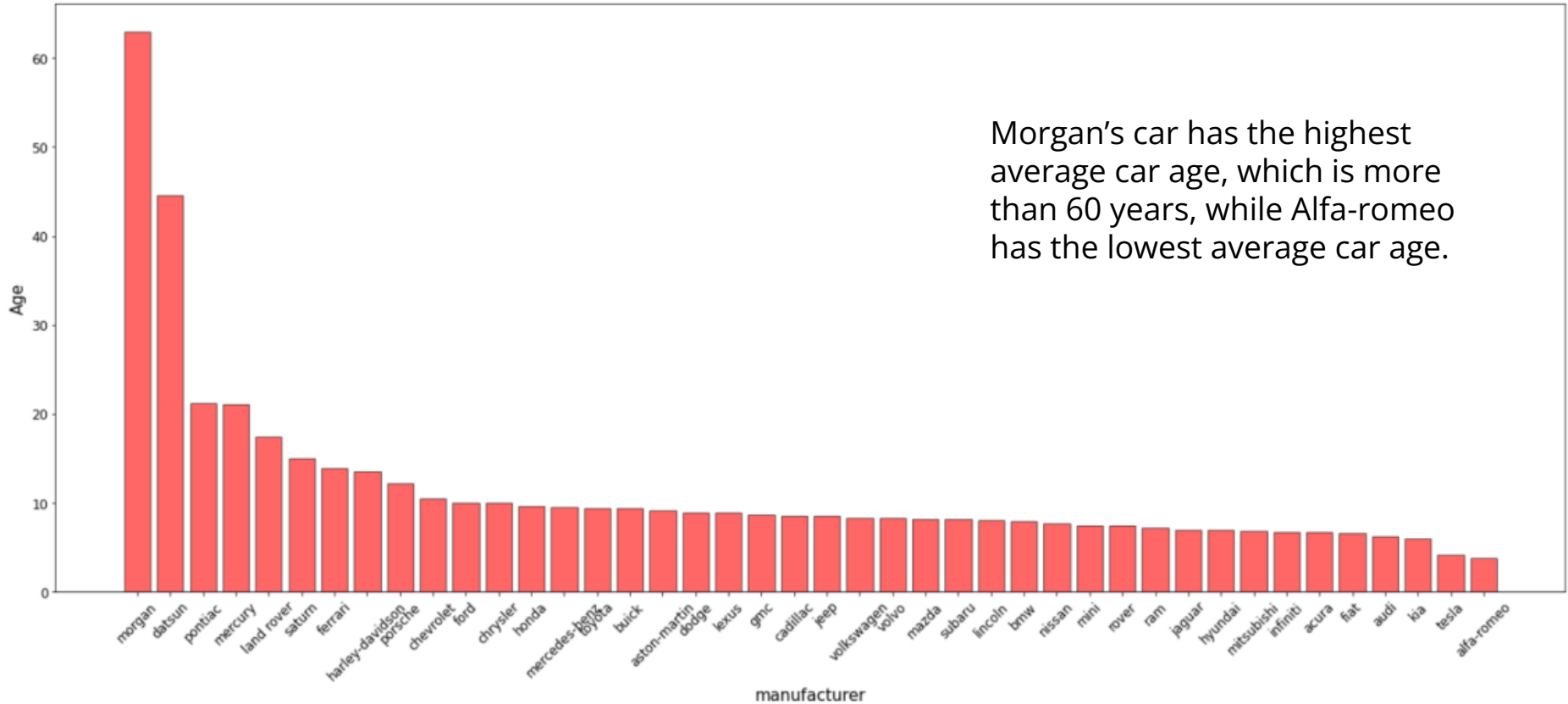


South Dakota has the highest average for car age, and Delaware has slightly lower average for car age than South Dakota. West Virginia and Alaska have relatively lower average car age among all the states.

Average Car Age by State

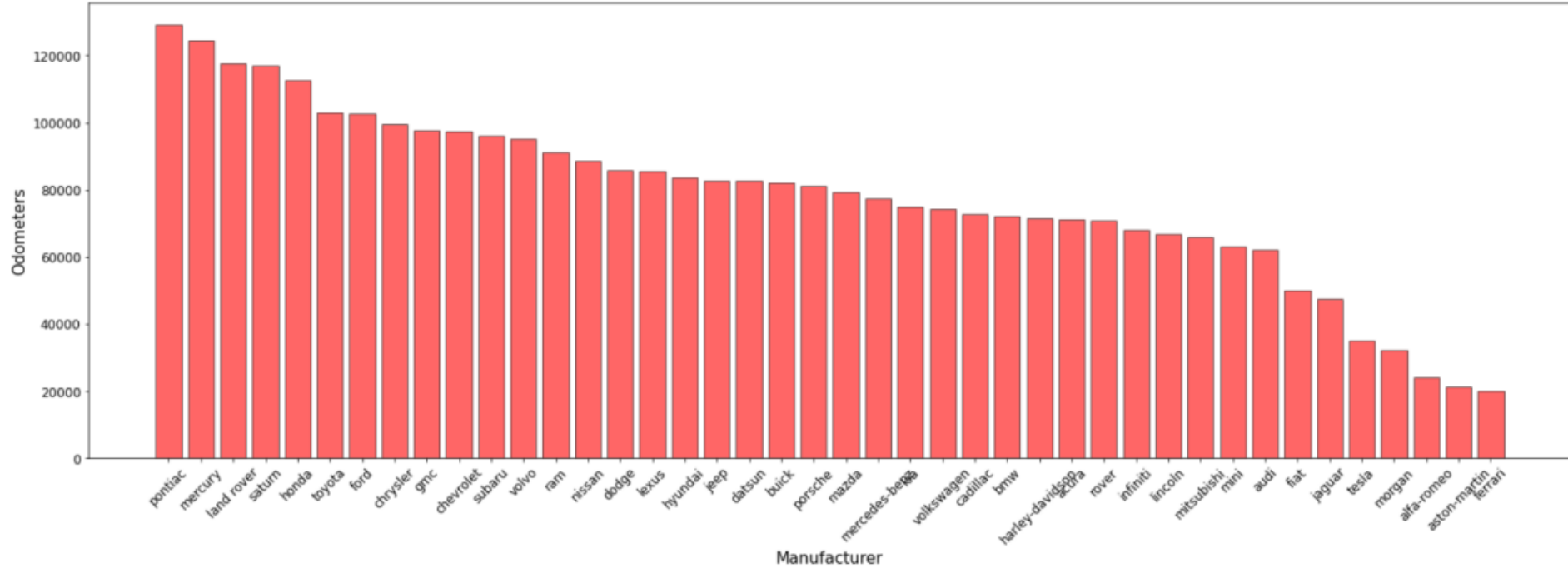


Average Car Age by manufacturer



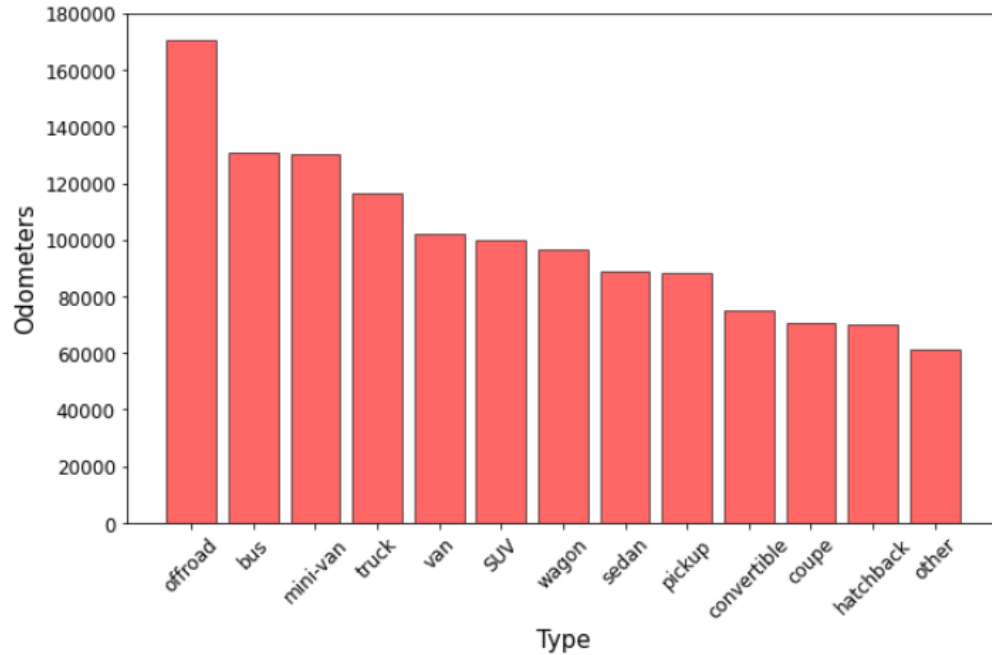
Morgan's car has the highest average car age, which is more than 60 years, while Alfa-romeo has the lowest average car age.

Average Odometers by Manufacturer



On average, Pontiac has the highest odometers, and ferrari has the lowest odometers. Most the other brands has an average odometer between 60,000 and 100,000 miles.

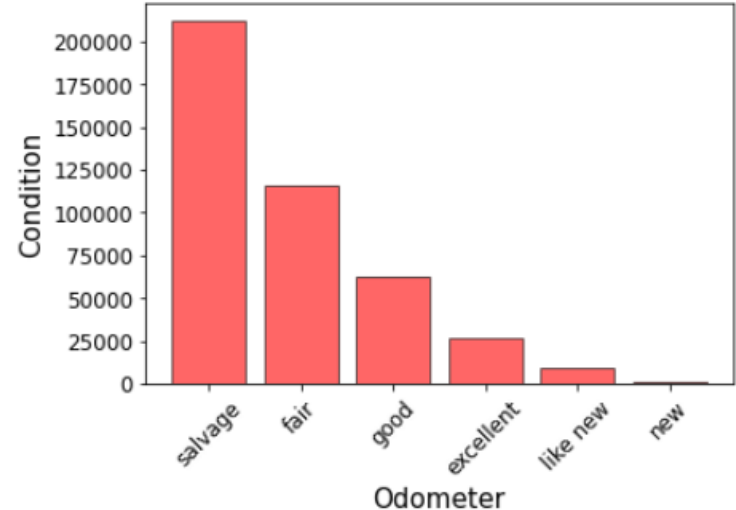
Average Odometers by Type



Salvage cars has higher odometers than other conditions, which makes sense as it is closer to its usage life.

Offroad has the highest average odometers, and other has the lowest average odometers.

Average Odometers by Condition



Research Question 1

How does the landform of a
location affect people's choice
of car?





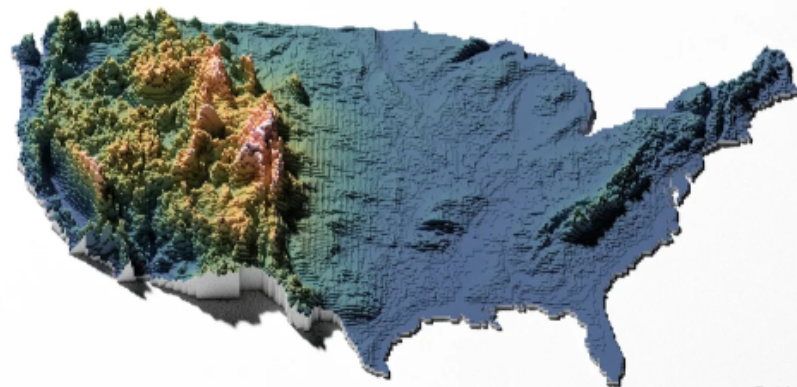
Mountain States Vs. Plain States

Mountain_states:

co, nm, mo, ar, wy, mt, ut, id, az, wa, or, nv, ca,
wv, pa, tn, vt, nh, me

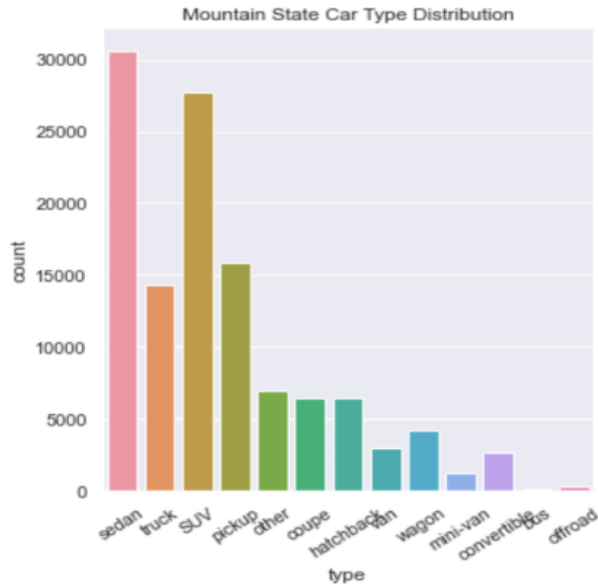
Slice into two data frames

Mountain States and Plain States

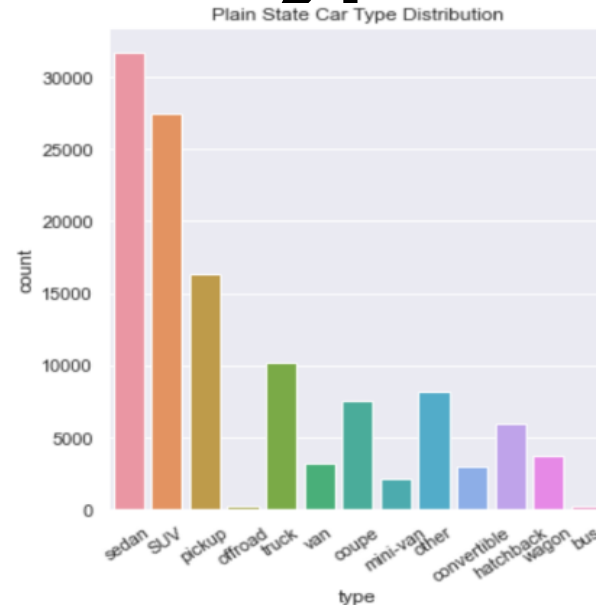


@costats1

Comparison by Car Type



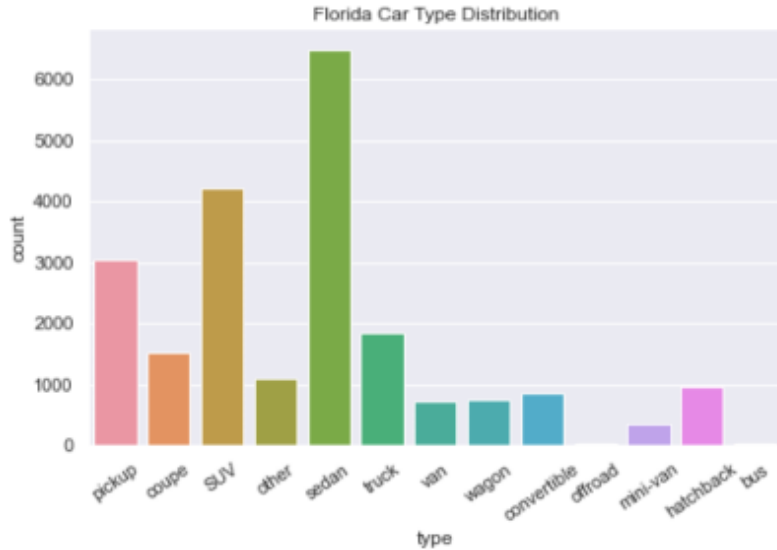
Mountain State



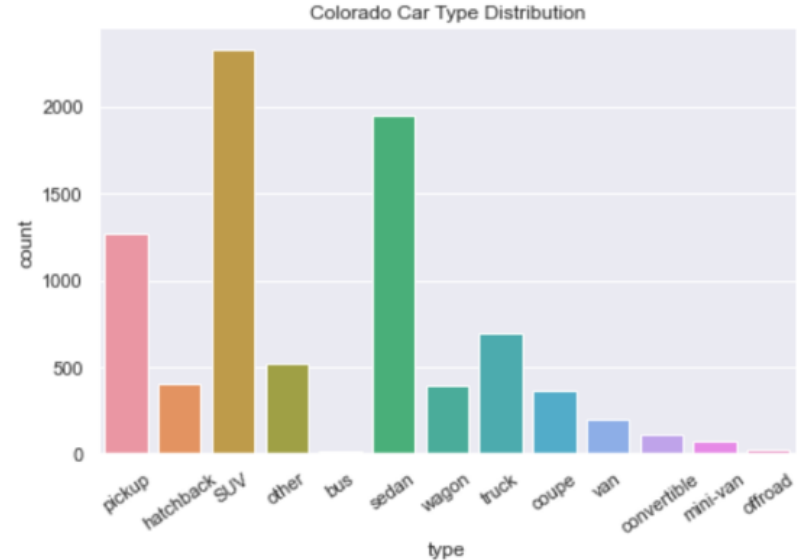
Plain State

Originally, We assume that in mountain states, SUV would be a more popular choice. But when comparing the most popular car types among mountain state and plain state, we found the top three are both sedan, SUV and pick up ranking 1, 2, 3. It seems that landform is not a major affecting factor when people are choosing car type.

Sample State Comparison



Florida



Colorado

To further testing our hypothesis, we pick Colorado and Florida as the two most representative sample state. In Colorado, where mountain is the major landform, we found SUV as the most popular car type. And in Florida, where plain is the major landform, the most popular car type is sedan. This result is inline with our initial hypothesis. We reached the conclusion that only when the landform is significantly different, will affect the car type choice.

Research Question 2

How does the weather of a
location affect people's choice
of car?



Snow States Vs. Non-Snow States

We define snow states as states have the average snow > 30 inches

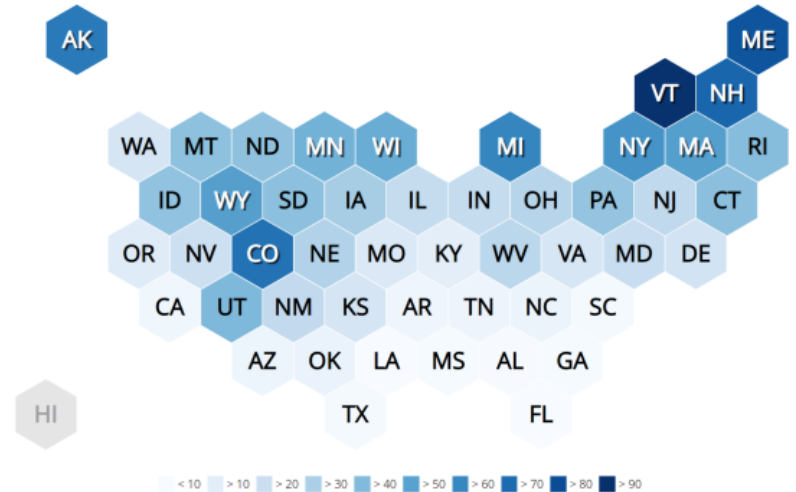
Snow States:

vt me nh co ak mi ny ma wy wi mn ut ri ct
sd mt nd id pa ia

Slice into two data frames

Snow States and Non-Snow States

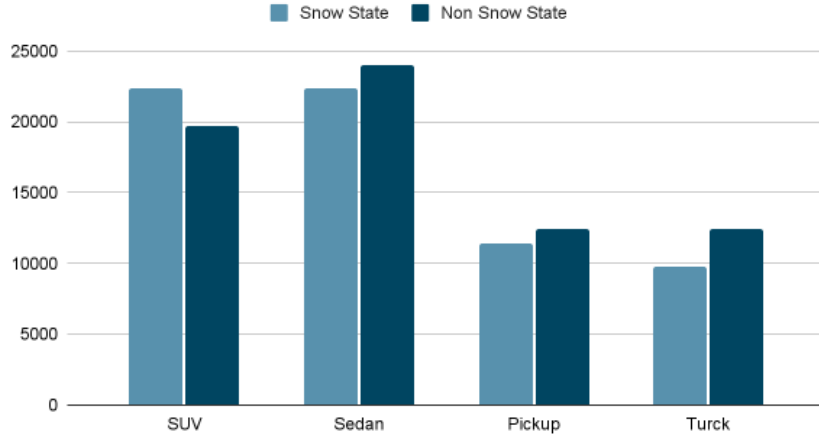
Snowiest States 2021



Source: <https://worldpopulationreview.com/state-rankings/snowiest-states>

Comparison

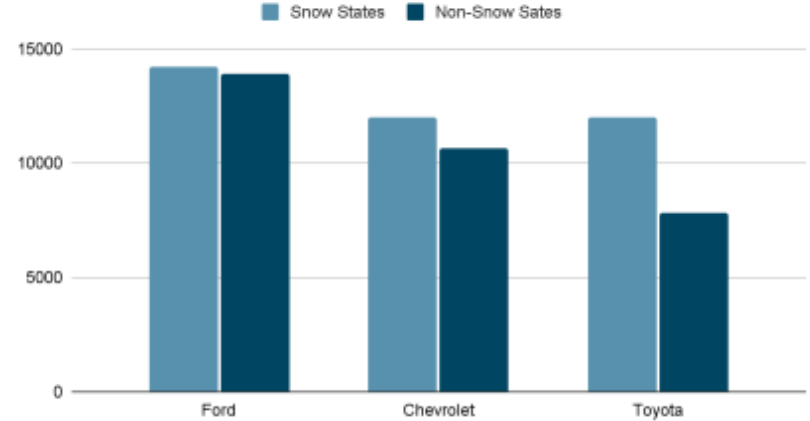
Snow State and Non Snow State



Car Type

In snow states, SUV is the most popular car type. In other states, sedan is the most popular

Manufacturers by States

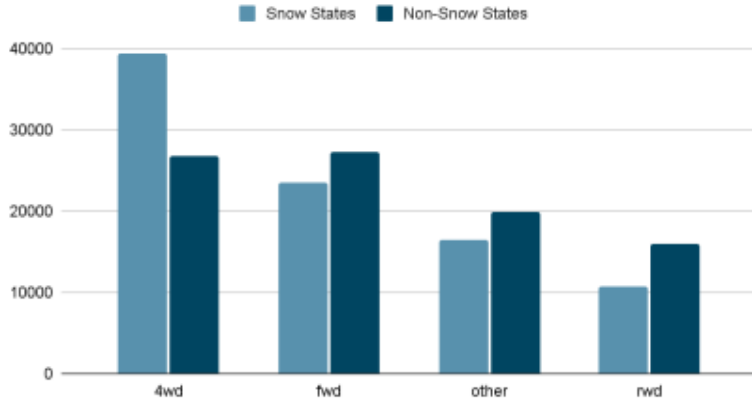


Manufacturer

Both snow states and non snow states has the same preference of car manufacturer

Comparison

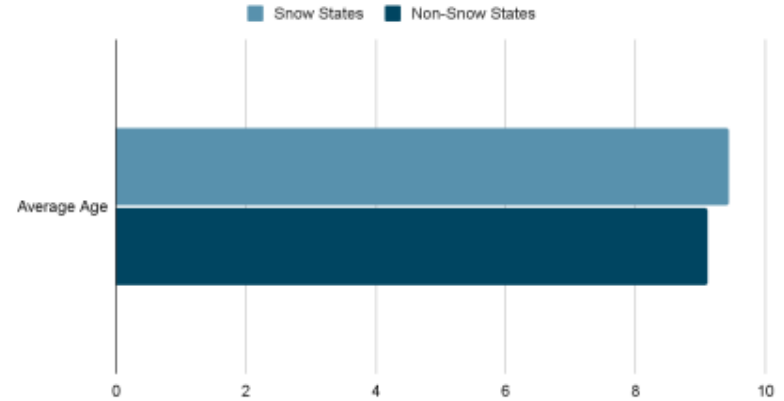
drive comparision



Drive

Forward-Drive Cars are more popular in non snow states and snow states prefer four-wheel drive

age comparison



Age

In non-snow states, people don't mind to driver older cars, compared to snow states



**Thank
you!**

