Price Prediction for Used Cars

A Microservice to Help You Predict Your Car's Worth

Aishwarya Padmanabha & Nissar Kotadia
Northeastern University

Agenda

- Problem Statement Overview
- All about the Dataset
  - Choice of Dataset
  - Exploratory Data Analysis
- Requirements
- Design & Implementation (showing in jupyter notebook)
  - Data Understanding
  - Data Pre Processing
  - Regression
  - Evaluating the Model
  - Save the Model
  - User Interface
- Evaluation
- Results
- Conclusion

Problem Statement Overview

Problem Statement Overview

Problem Definition

Given some specifications of a car, provide a rough estimate of its current worth

Why we chose this problem

- Based on dataset - LOTS of features and LOTS of data points to analyse
- Can be used in the real world by many people

Requirements

Requirements

- python v(3.6.9)
- scikit-learn v(0.22.2.post1)
- numpy v(1.18.2)
- pandas v(1.0.3)
- joblib v(0.14.1)
- matplotlib v(3.2.1)
- connexion v(2.6.0)
- seaborn

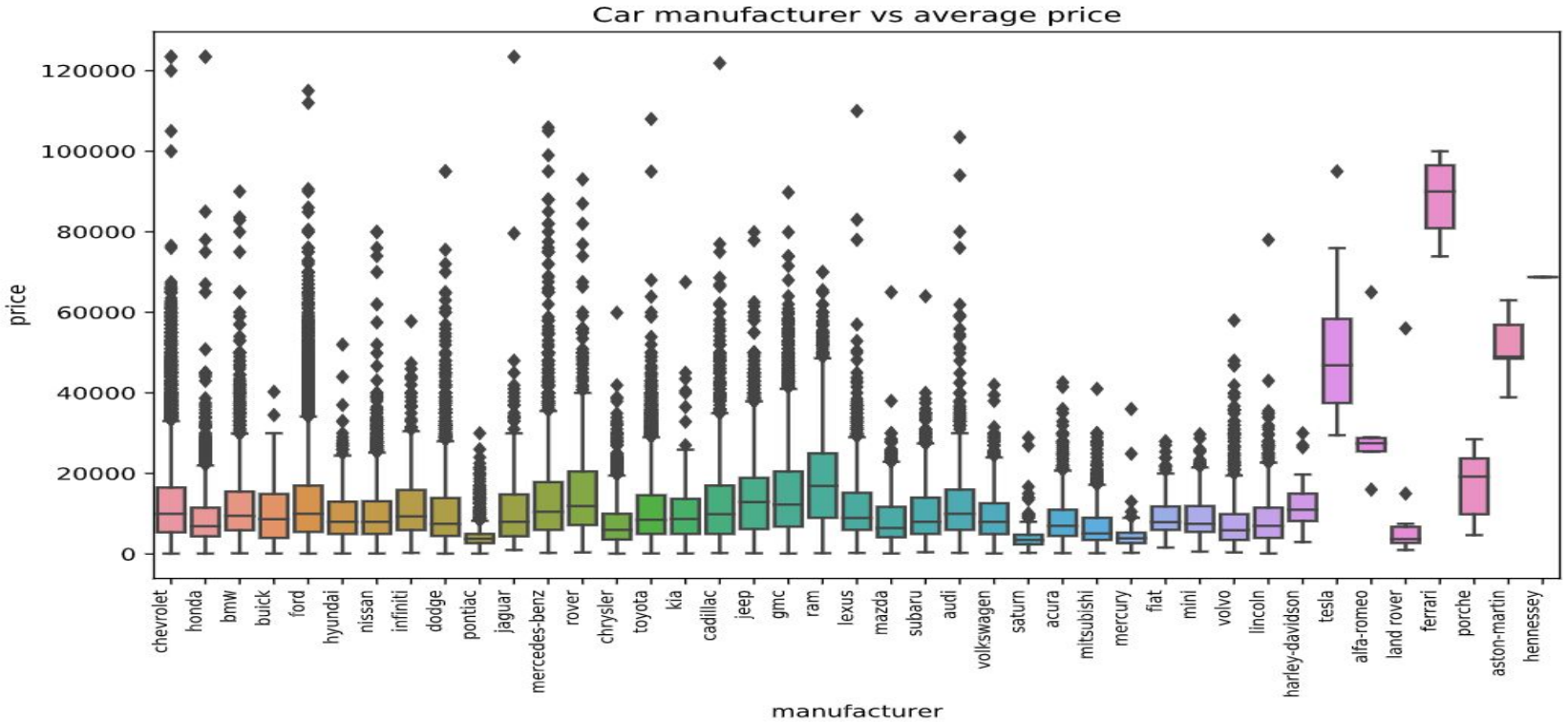
All about the Dataset

All about the Dataset

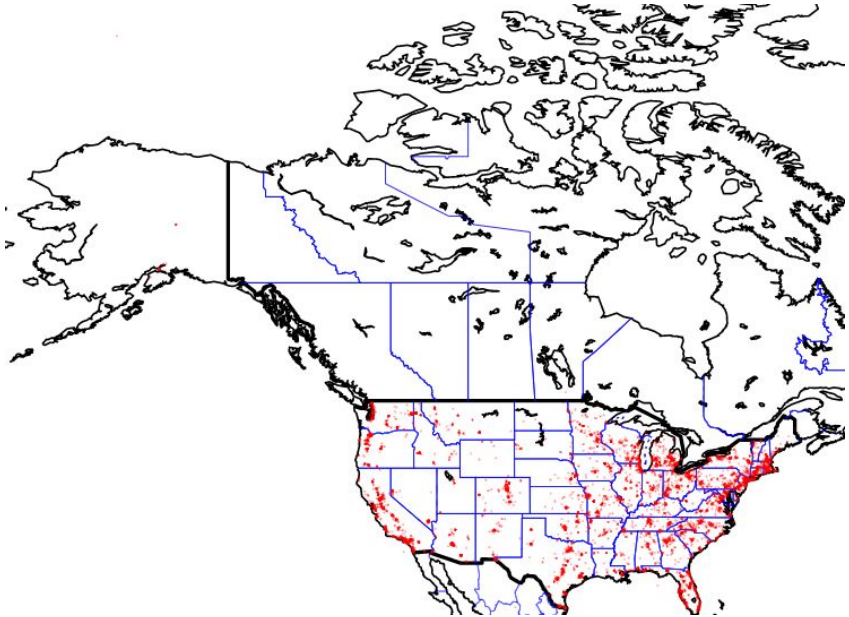
Choice of Dataset:

- What we wanted in our dataset:
  - Multiple data points to achieve higher accuracy check for trends/patterns in the data
  - Multiple features to make it more robust.
  - Basic attributes must exist (in our case, price of a car and at minimum, the name of the car)
- How the used cars dataset fit our needs:
  - Basic attributes of cars and it's price exist
  - 25 columns (id, url, region, region_url, price, year, manufacturer,model, condition, cylinders, fuel, odometer, title_status,transmission, vin, drive, size, type, paint_color,image_url, description, county, state, lat, long)

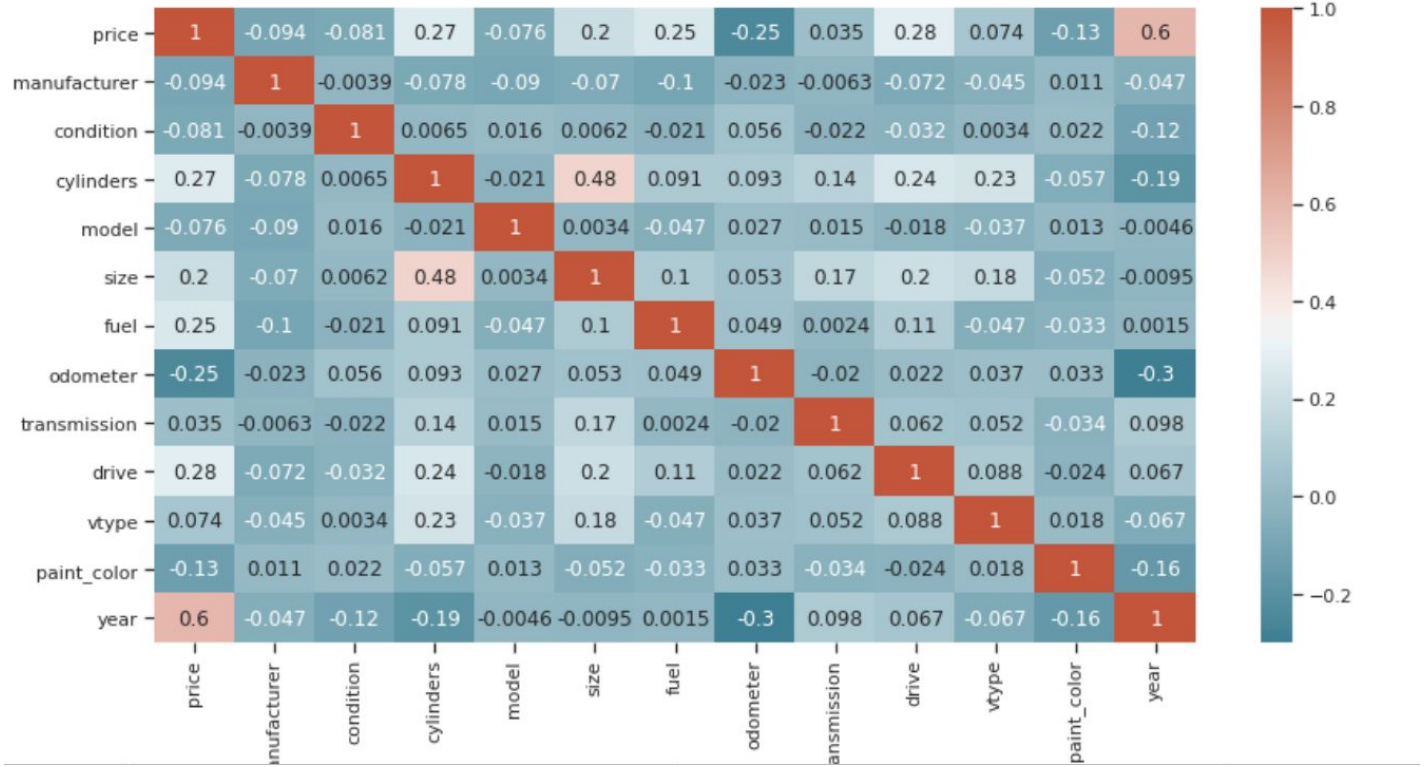
Link: https://www.kaggle.com/austinreese/craigslist-carstrucks-dat


Car manufacturer vs average price
price
0
20000
40000
60000
80000
100000
120000
chevrolet
honda
bmw
buick
ford
hyundai
nissan
infiniti
dodge
pontiac
jaguar
mercedes-benz
rover
chrysler
toyota
kia
cadillac
jeep
gmc
ram
lexus
mazda
subaru
audi
volkswagen
saturn
acura
mitsubishi
mercury
fiat
mini
volvo
lincoln
harley-davidson
tesla
alfa-romeo
land rover
ferrari
porche
aston-martin
hennessey
manufacturer

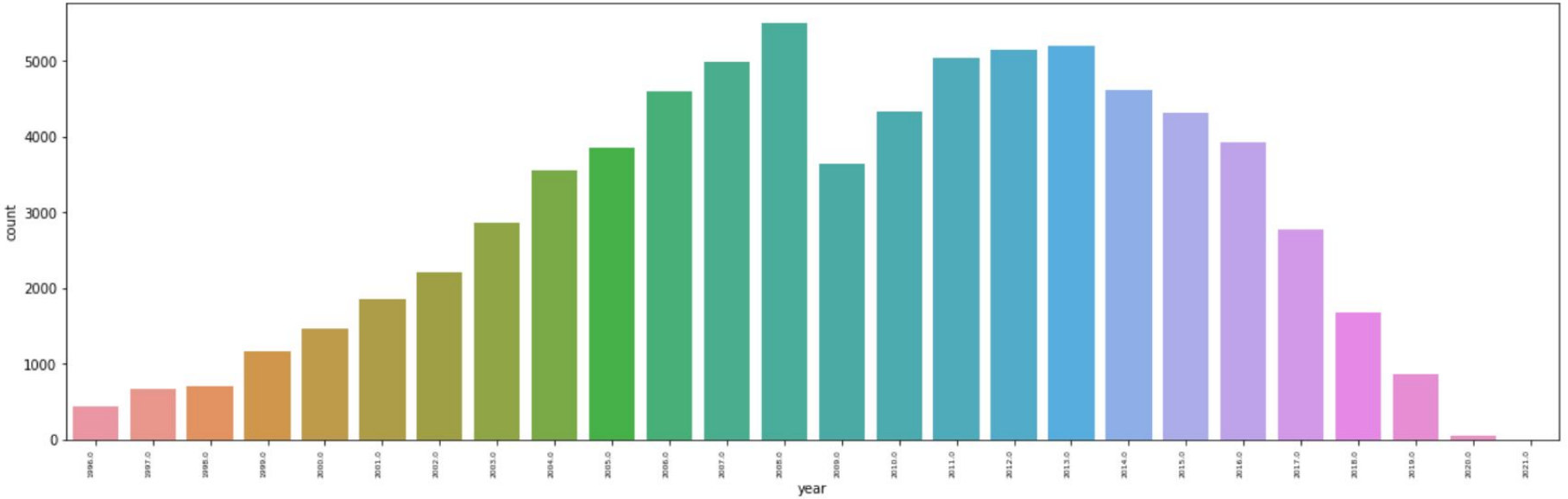
Origins of various data points



Correlation between different features



Number of instances across different years



Design and Implementation

(switch to jupyter-lab..)

User Interface

Car Price Prediction API

An API that can predicts the price of used cars

default Show/Hide | List Operations | Expand Operations

POST /predict — Entrypoint to our prediction function

Parameters

Parameter	Value	Description	Parameter Type	Data Type
manufacturer	volkswagen	company that produced the vehicle	query	string
model	golf r	make of the vehicle	formData	string
condition	excellent	current state of the vehicle	query	string
cylinders	4 cylinders	number of cylinders	query	string
fuel	gas	type of fuel	query	string
odometer	23700	distance travelled by the vehicle	formData	string
transmission	manual	type of transmission system of the vehicle	query	string
drive	4wd	type of drive	query	string
size	compact	size of the vehicle	query	string
vtype	hatchback	configuration of the vehicle	query	string
paint_color	black	color of the vehicle body	query	string
year	2002	manufacturing year of the vehicle	formData	double

Response Messages

HTTP Status Code	Reason	Response Model	Head
200	Prediction performed successfully		
500	Unable to perform the prediction		

Try it out! Hide Response

Evaluation

Evaluation

R2 score (scores recorded are the highest we've seen for each regression algorithm after running it multiple times):

- RandomForestRegressor: 0.8556
- GradientBoostingRegressor: 0.8029
- LogisticRegression: 0.6560

Results

Results (Input)

Car Price Prediction API

An API that can predicts the price of used cars

default Show/Hide | List Operations | Expand Operations

POST /predict — Entrypoint to our prediction function

Parameters

Parameter	Value	Description	Parameter Type	Data Type
manufacturer	volkswagen	company that produced the vehicle	query	string
model	golf r	make of the vehicle	formData	string
condition	excellent	current state of the vehicle	query	string
cylinders	4 cylinders	number of cylinders	query	string
fuel	gas	type of fuel	query	string
odometer	23700	distance travelled by the vehicle	formData	string
transmission	manual	type of transmission system of the vehicle	query	string
drive	4wd	type of drive	query	string
size	compact	size of the vehicle	query	string
vtype	hatchback	configuration of the vehicle	query	string
paint_color	black	color of the vehicle body	query	string
year	2002	manufacturing year of the vehicle	formData	double

Response Messages

HTTP Status Code	Reason	Response Model	Head
200	Prediction performed successfully		
500	Unable to perform the prediction		

Try it out! Hide Response

Results (Output)

Response Body


```
{
  "prediction": "6462.475"
}
```