

USED CARS DATASET..

By : Soaad Aljafr.
Reem Alghamdi.



• Content..

- Interdiction
- Problem statement
- Data
- Data Cleaning
- Visualizations

Interdiction..

- 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 the world's largest collection of used vehicles for sale, dataset which includes every used vehicle entry within the United States on Craigslist. it contains most all relevant information that provides on car sales including columns like price, condition, manufacturer, latitude/longitude, and 18 other categories.

Problem statement:

- Knowing the sale price of the used cars is important for the sellers and buyers, whenever the price of the used car accrued price increased the proportion of selling used cars. Therefore, this project will suggest a machine Learning regression model that predicts the accrued price to sale the used car.

Data

- Used Cars Dataset
- Vehicles listings from Craigslist.org
- Rows is above 420000
- Columns is 26

Data Cleaning

- Read The Data from cvs
- Drop some colums
- Delet the Null values
- Fill Null values
- Change the object colums to numaric

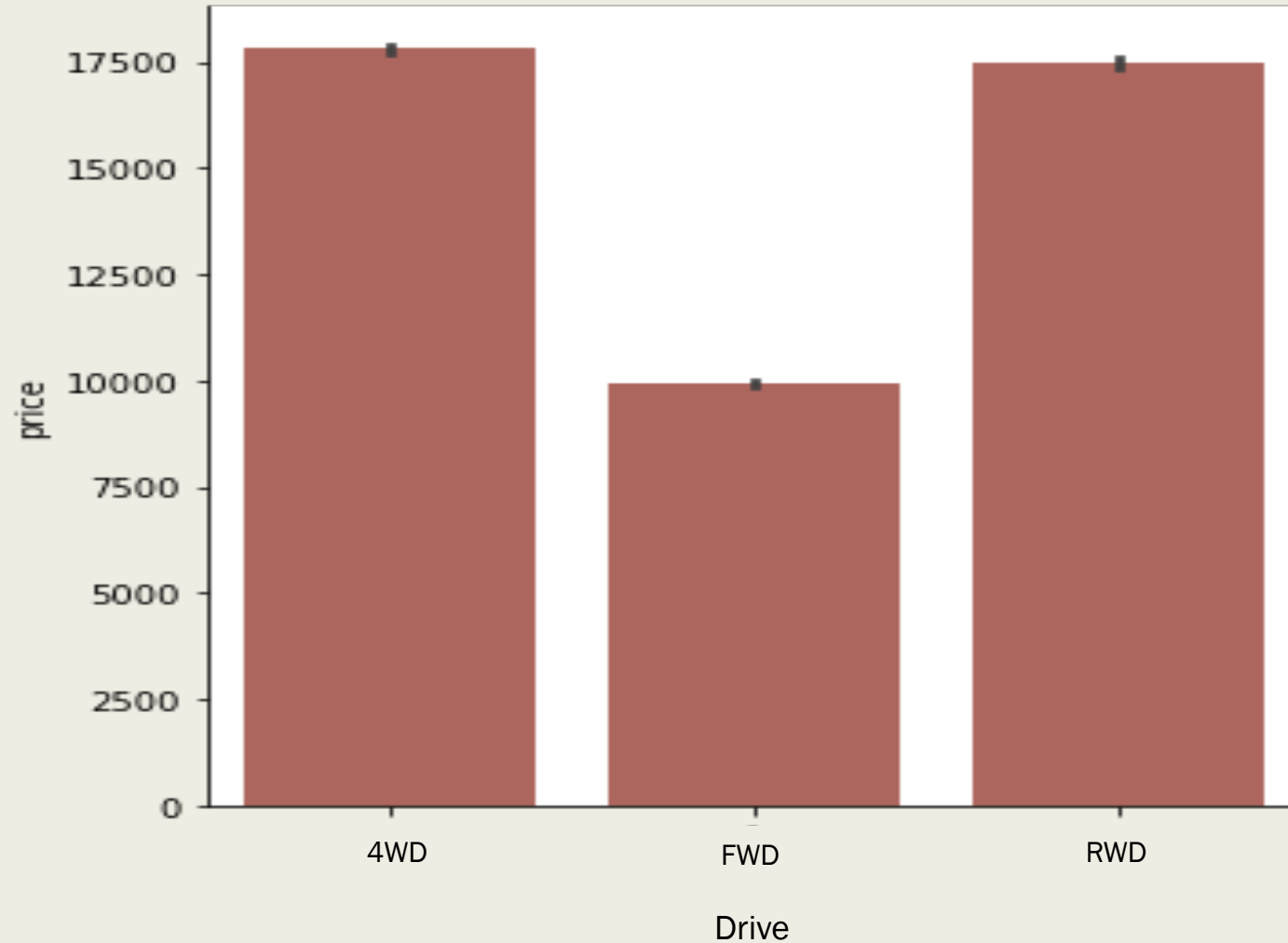
Num. null

	0
region	0
price	0
year	1205
manufacturer	17646
model	5277
condition	174104
cylinders	177678
fuel	3013
odometer	4400
title_status	8242
transmission	2556
drive	130567
size	306361
type	92858
paint_color	130203
county	426880
state	0
posting_date	68

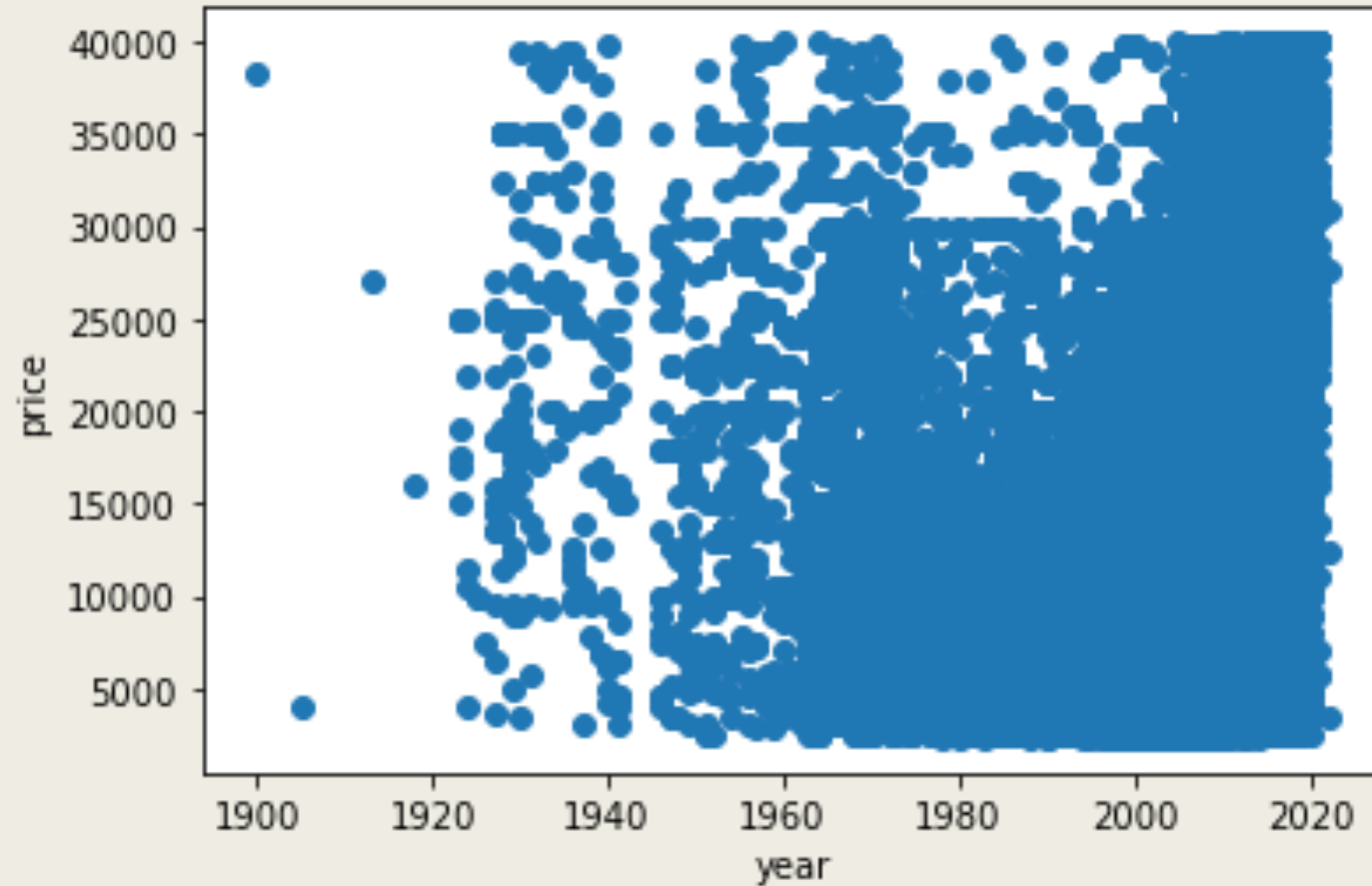
Object describe

	COUNT	UNIQUE	TOP	FREQ
region	426880	404	columbus	3608
manufacturer	409234	42	ford	70985
model	421603	29667	f-150	8009
condition	252776	6	good	121456
cylinders	249202	8	6 cylinders	94169
fuel	423867	5	gas	356209
title_status	418638	6	clean	405117
transmission	424324	3	automatic	336524
drive	296313	3	4wd	131904
type	334022	13	sedan	87056
paint_color	296677	12	white	79285
state	426880	51	ca	50614
posting_date	426812	381536	2021-04-23T22:13:05-0400	12

Bar plot between type and price



Scatler Years and Price



Data Correlation



Model and Result

R ²	Training	Validation	Test
Liner Regression	0.5593	0.5596	0.570
Polynomial	0.7373	0.852	0.87

Thanks..