Introduction

The repository is a project of using the The dataset "Used Cars Dataset" (https://www.kaggle.com/austinreese/craigslist-carstrucks-data) to predict a linear relationship between "odometer of vehicles", "verhicle drive types(front-wheel drive etc)", "vehicle types(SUV etc)", "paint color of vehicles" and the price of the used cars in United States.

The notebook "car_price_predict.ipynb" provides all the codes, plots, comments, and summary of the data science process of discovering the relationship between the listed price and our predictors. The video in this repository walks you through the notebook by providing detailed explanation of author's ideas.

The Dataset containing collections of used vehicles for sale on Craigslist. The dataset includes every used vehicle entry within the United States on Craigslist and is updated every few months. The dataset I am using is last updated in January 2020. The dataset is about 1.42 GB.

Some Statistics and plots from the notebook

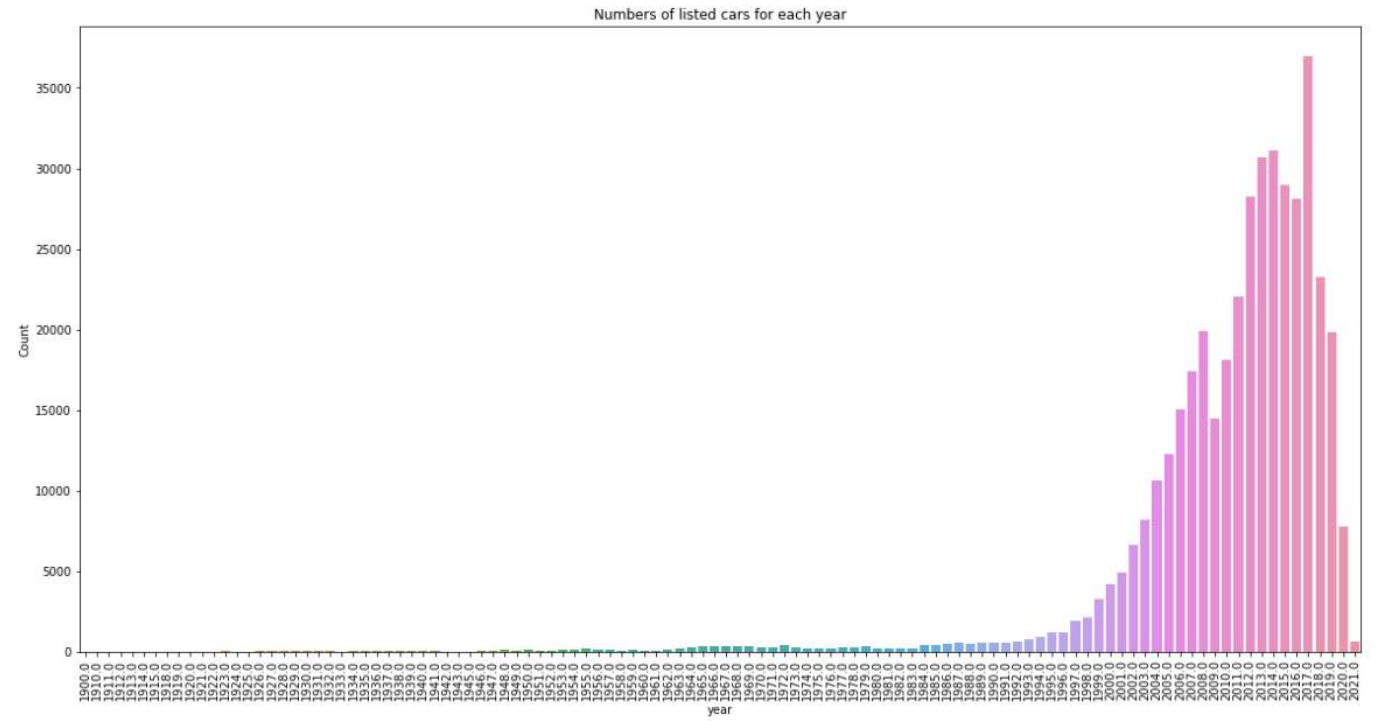
Here are 3 plots from the notebook to give you an idea of how I work with the dataset

Here is a glance at the data

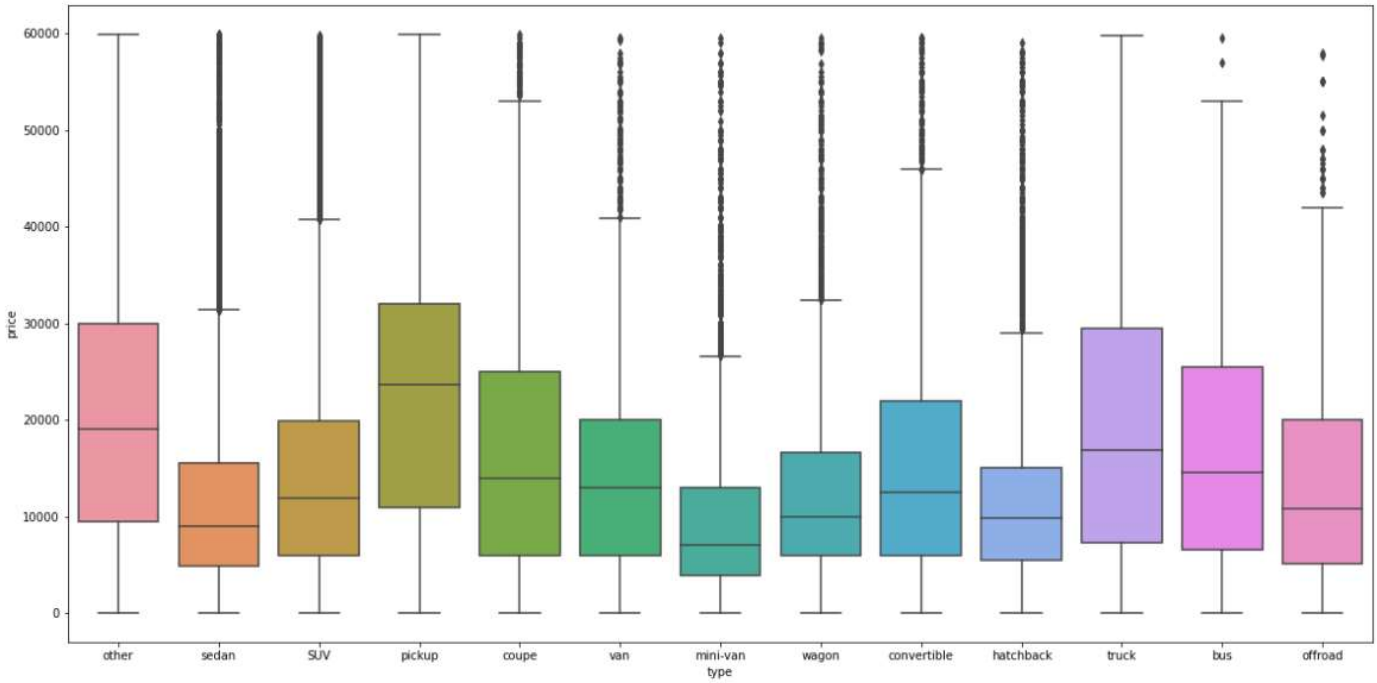
	id	url	region	region_url	price	year	manufacturer	model	condition	...	drive	size	type	paint_color	image_url	description	state
	7240372487	https://auburn.craigslist.org/ctd/d/auburn-uni...	auburn	https://auburn.craigslist.org	35990	2010.0	chevrolet	corvette grand sport	good	...	rwd	NaN	other	NaN	https://images.craigslist.org/00N0N_ipkbHVZYf4...	Carvana is the safer way to buy a car During t...	al
	7240309422	https://auburn.craigslist.org/cto/d/auburn-201...	auburn	https://auburn.craigslist.org	7500	2014.0	hyundai	sonata	excellent	...	fwd	NaN	sedan	NaN	https://images.craigslist.org/00s0s_gBHYmJ5o7y...	I'll move to another city and try to sell my c...	al
	7240224296	https://auburn.craigslist.org/cto/d/auburn-200...	auburn	https://auburn.craigslist.org	4900	2006.0	bmw	x3 3.0i	good	...	NaN	NaN	SUV	blue	https://images.craigslist.org/00B0B_5zgEGWPOrt...	Clean 2006 BMW X3 3.0i. Beautiful and rare Bl...	al
	7240103965	https://auburn.craigslist.org/cto/d/lanett-tru...	auburn	https://auburn.craigslist.org	2000	1974.0	chevrolet	c-10	good	...	rwd	full-size	pickup	blue	https://images.craigslist.org/00M0M_6o7KcDpArw...	1974 chev. truck (LONG BED) NEW starter front ...	al
	7239983776	https://auburn.craigslist.org/cto/d/auburn-200...	auburn	https://auburn.craigslist.org	19500	2005.0	ford	f350 lariat	excellent	...	4wd	full-size	pickup	blue	https://images.craigslist.org/00p0p_b95I1EgUfl...	2005 Ford F350 Lariat (Bullet Proofed). This t...	al

olumns

A plot of number of listed cars for each year from the dataset



Car type and their corresponding price distribution.



Data Collection Process

The notebook I implemented is a kaggle notebook and the dataset is already available on kaggle. I

If you have trouble opening the notebook or other issues, please send an email to wangr@lafayette.edu