



FINDING REAL ESTATE DEALS (& PEACE)



NF_project_1 - David - 2020 / 09 / 17



AIM OF MY EDA





TABLE OF CONTENTS

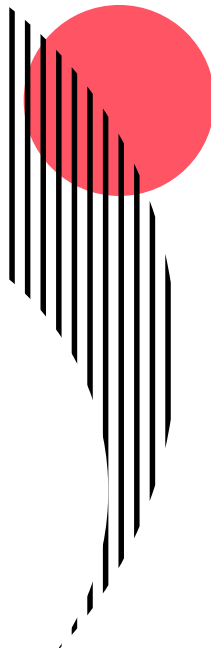
01 **INTRO**
What I have done.

03 **PRICE PREDICTION**
Will we get rich?

02 **EXPLORATION**
Interesting findings.

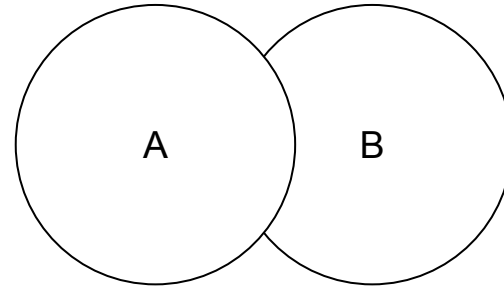
04 **NEXT STEPS**
What's coming next?

The dataset(s)



A
King County
House prices

B
King County
Sheriff Office - Incidents



left join on zip code

Tools & Packages

Data Handling



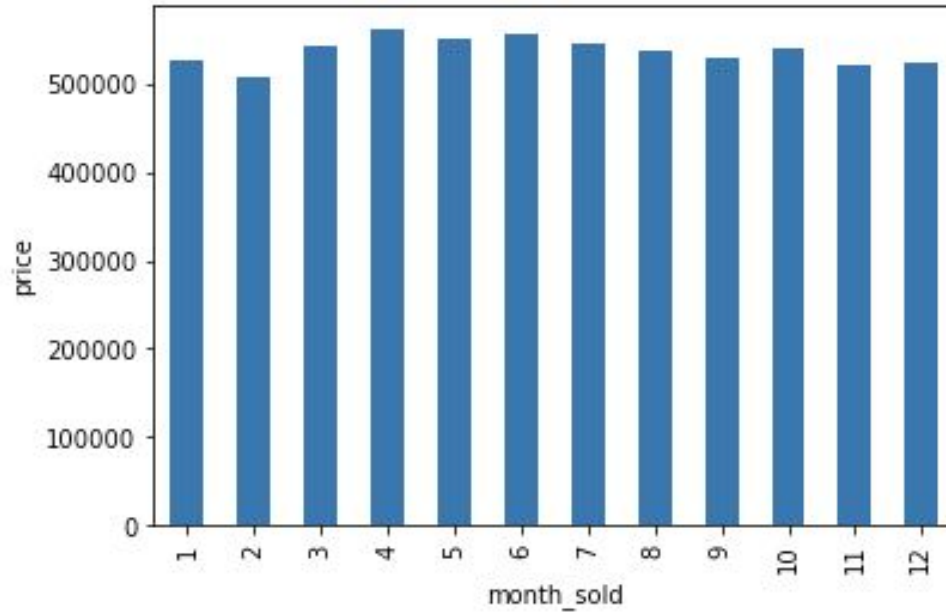
Data Visualization



ML

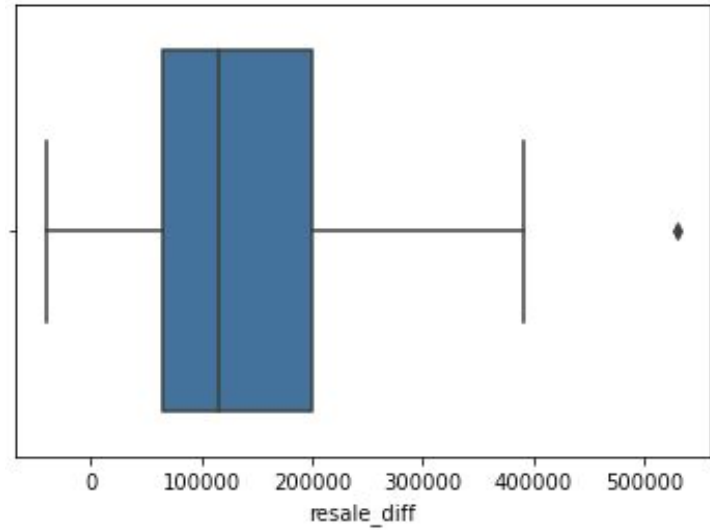
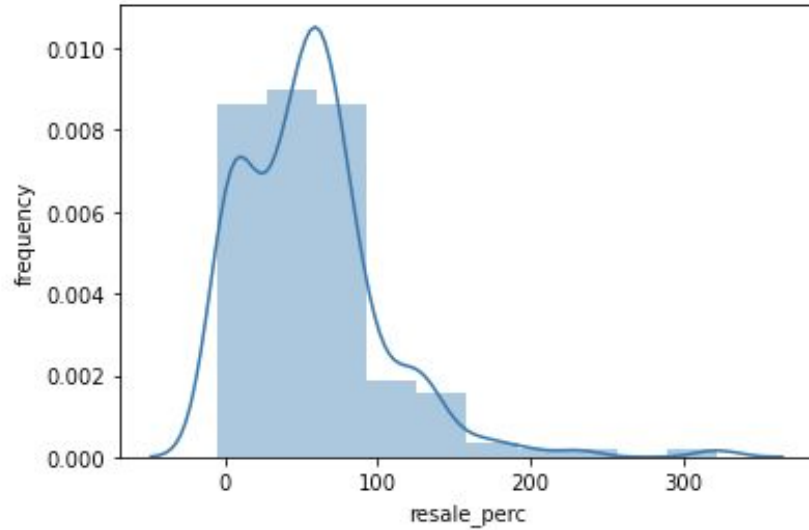


Prices per month



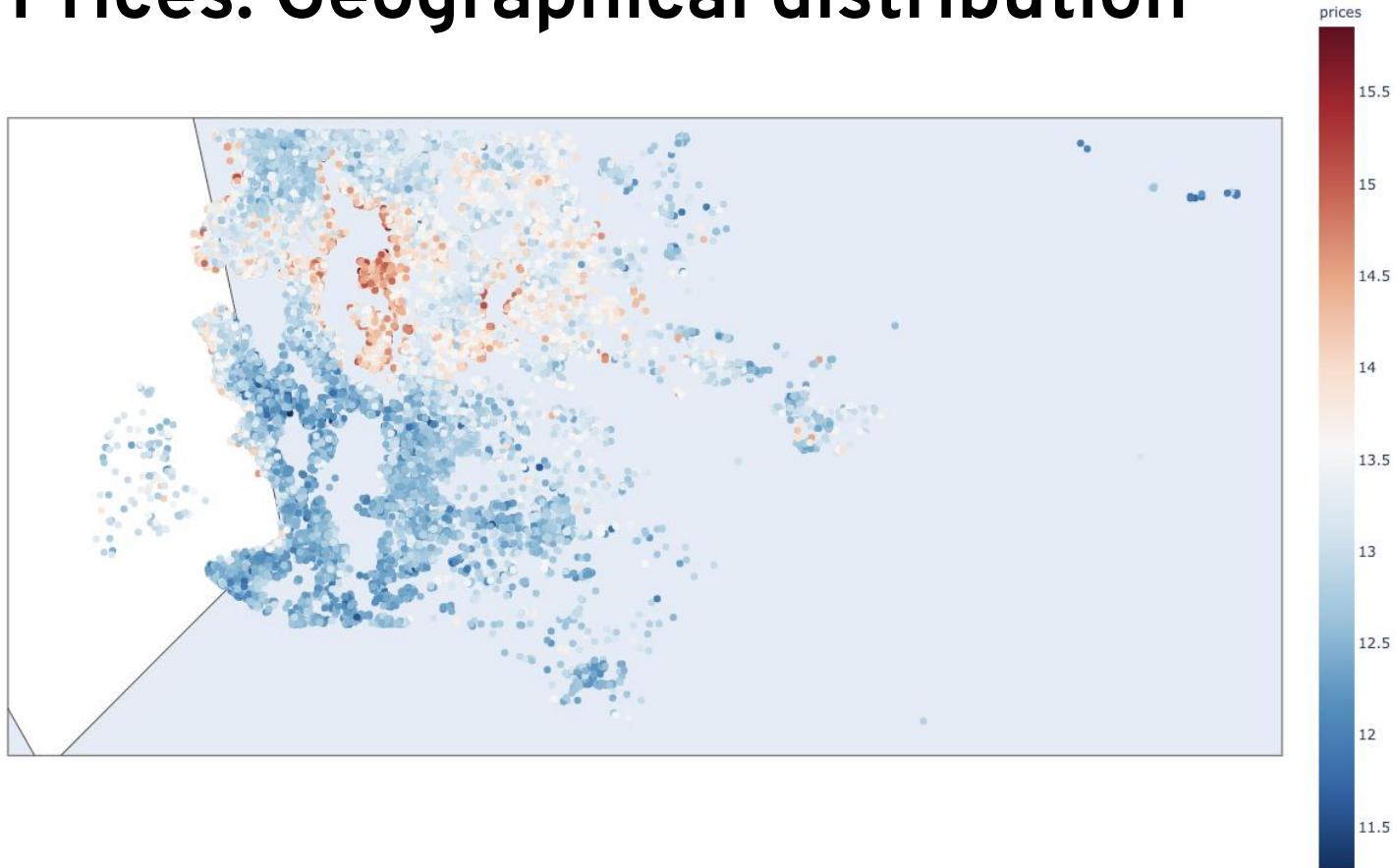
month_sold	count	mean
1	978	525963
2	1247	508520
3	1875	544058
4	2229	562216
5	2414	550850
6	2178	557534
7	2211	544892
8	1939	536655
9	1771	529724
10	1876	539439
11	1409	522360
12	1470	524800

Resales

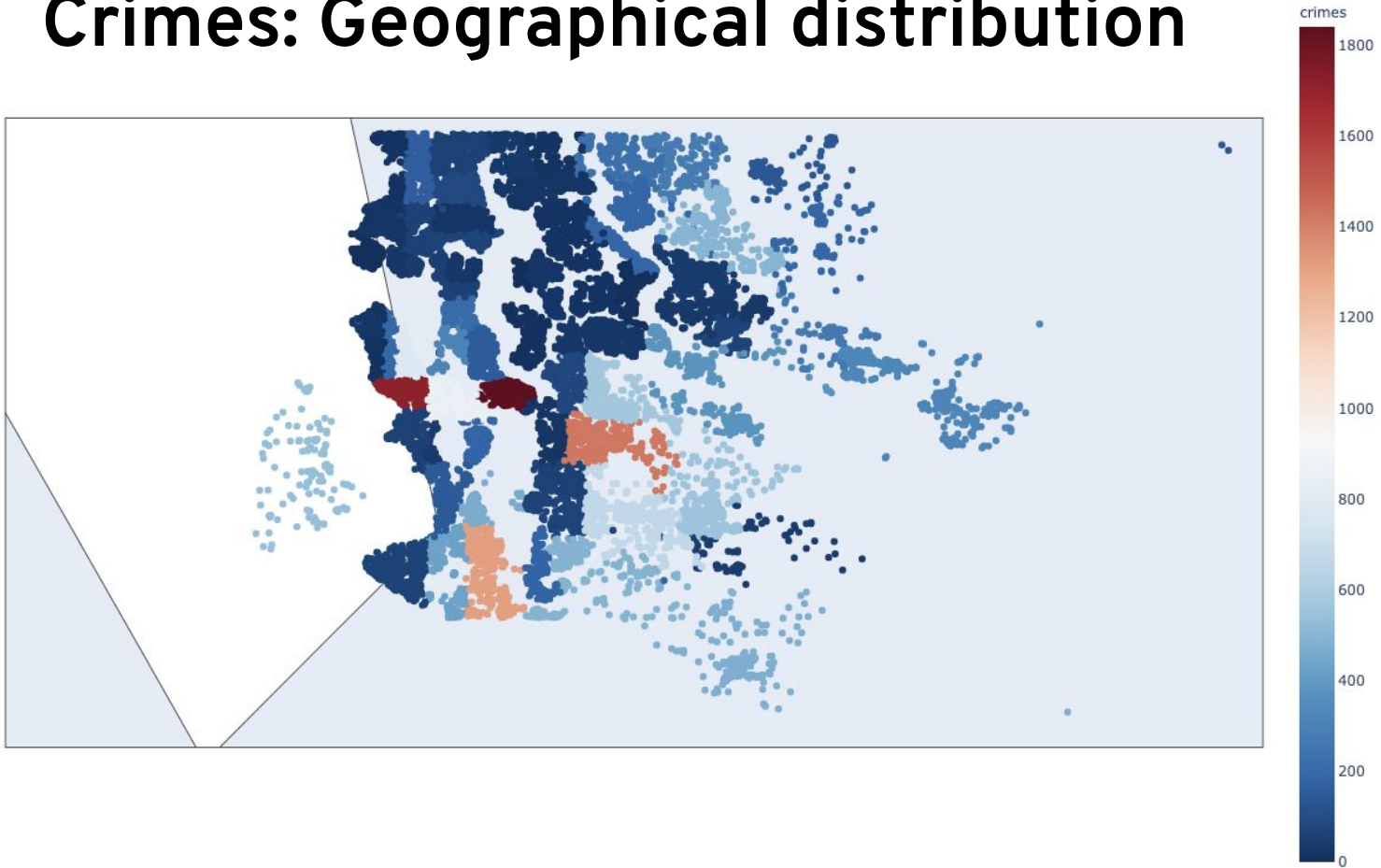


	count	mean	std	min	5%	25%	50%	75%	max
resale_diff	177.0	135796.0	100597.0	-40000.0	4000.0	65000.0	115450.0	199000.0	529950.0
resale_perc	177.0	56.0	47.0	-5.0	1.0	23.0	54.0	74.0	322.0

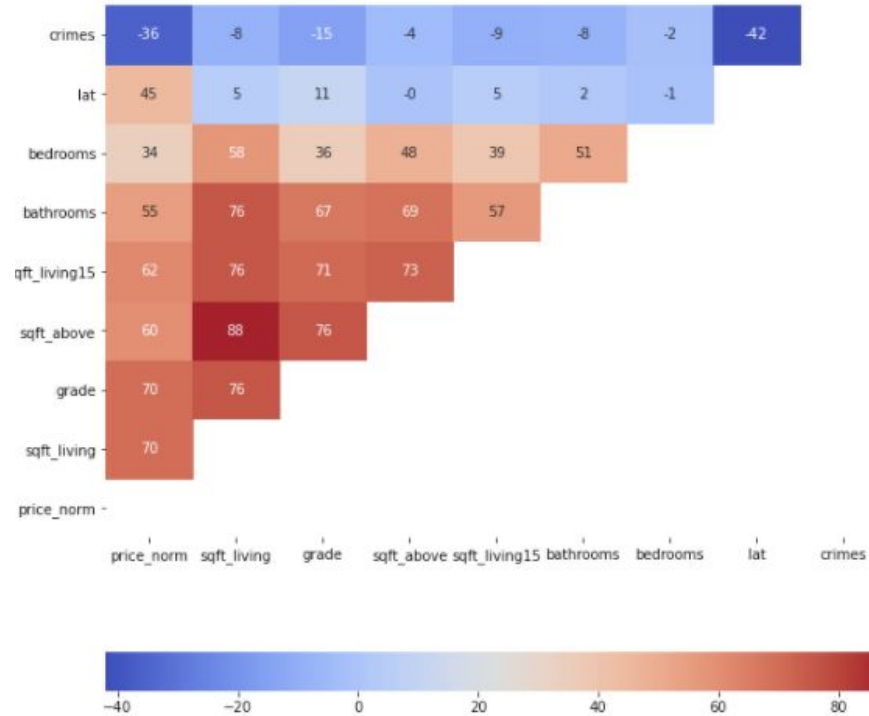
Prices: Geographical distribution



Crimes: Geographical distribution



Correlations in focus



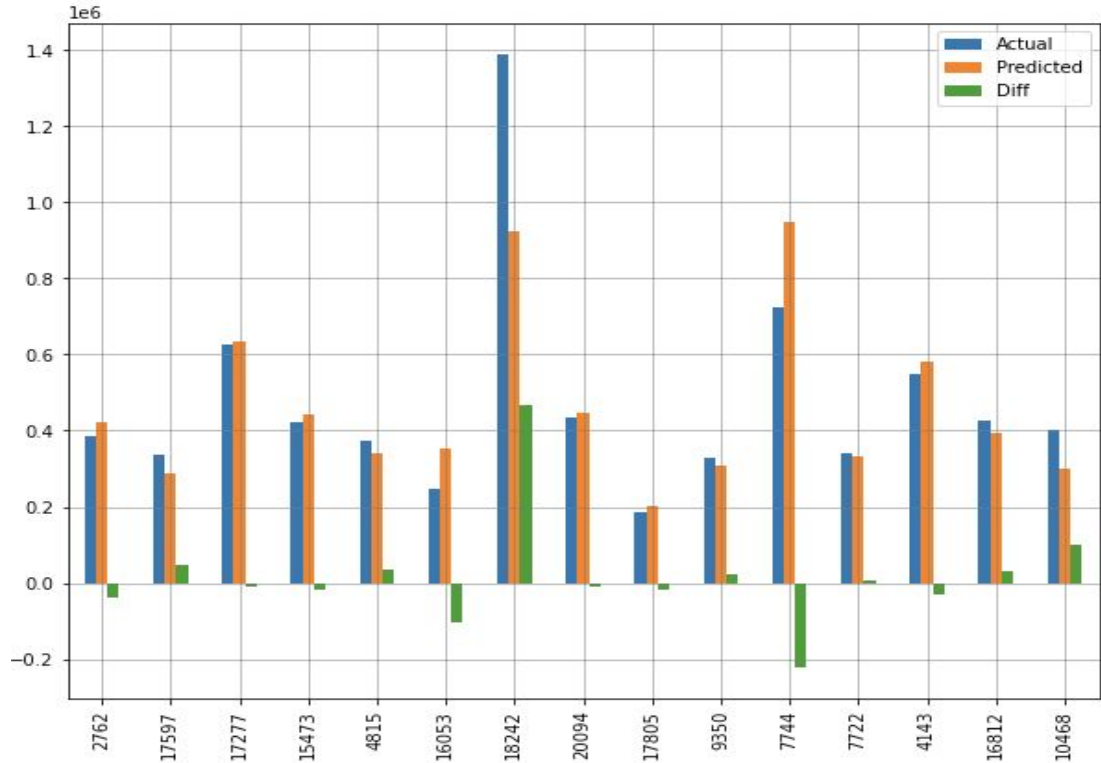


MULTIVAR LINEAR REGRESSION

Feature	Coefficient
sqrt_living_room	0.481
bathrooms	0.022
bedrooms	-0.027
condition	0.094
grade	0.168
lat	1.302

PREDICTION RESULTS

- R2 score:
0.721
- MSE score:
0.075





NEXT STEPS

- Taking a rest break
- Optimize the data cleaning
- Clustering crime data and investigate date



THANK YOU!

https://github.com/dave-py/eda_real-estate.git