HOUSE PRICE PREDICTION

Group III

HOUSE PRICE PREDICTION

— House prices increase every year, so there is a need for a system to predict house prices in the future.

House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.

TARGET MARKET

- **Family buyers**

Family orientated buyers have very specific requirements when it comes to buying a house. Deciding factors for these buyers typically include proximity to good schools and all necessary amenities such as hospitals.

"We will use alternative data that include schools, hospitals, food facilities and police station within a radius of 5 km."

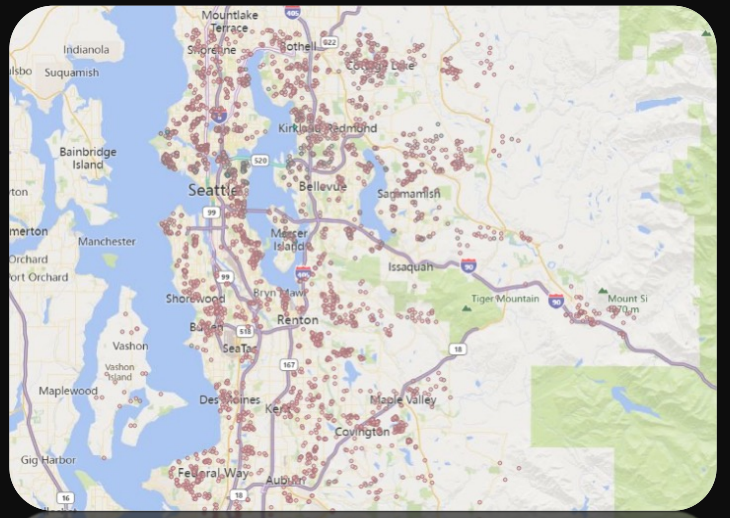
OBJECTIVE

- Predict the sale price for each house.
- Find insight of features in data
- Minimize the difference between predicted and actual rating (RMSE/MSE)
- Find the best model for Prediction

DATA

MAIN DATA

House Sales in “King County Washington State, USA” dataset contains house sale prices. It includes homes sold between May 2014 and May 2015.



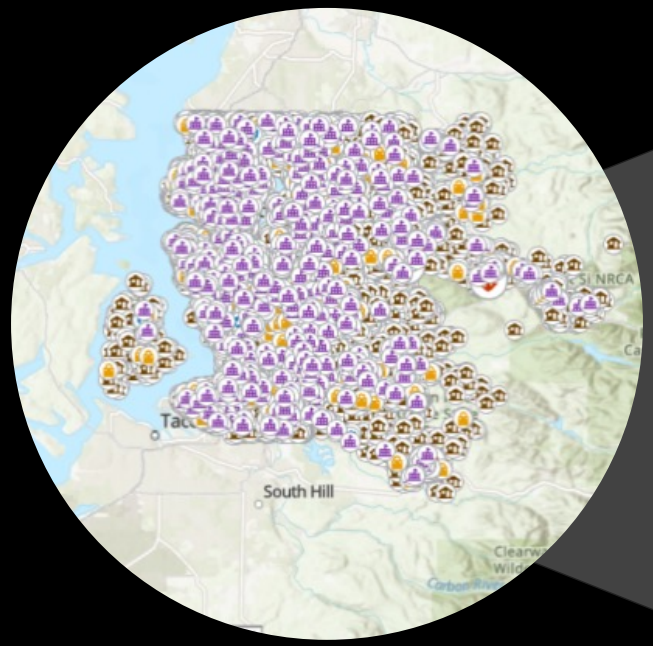
ALTERNATIVE DATA

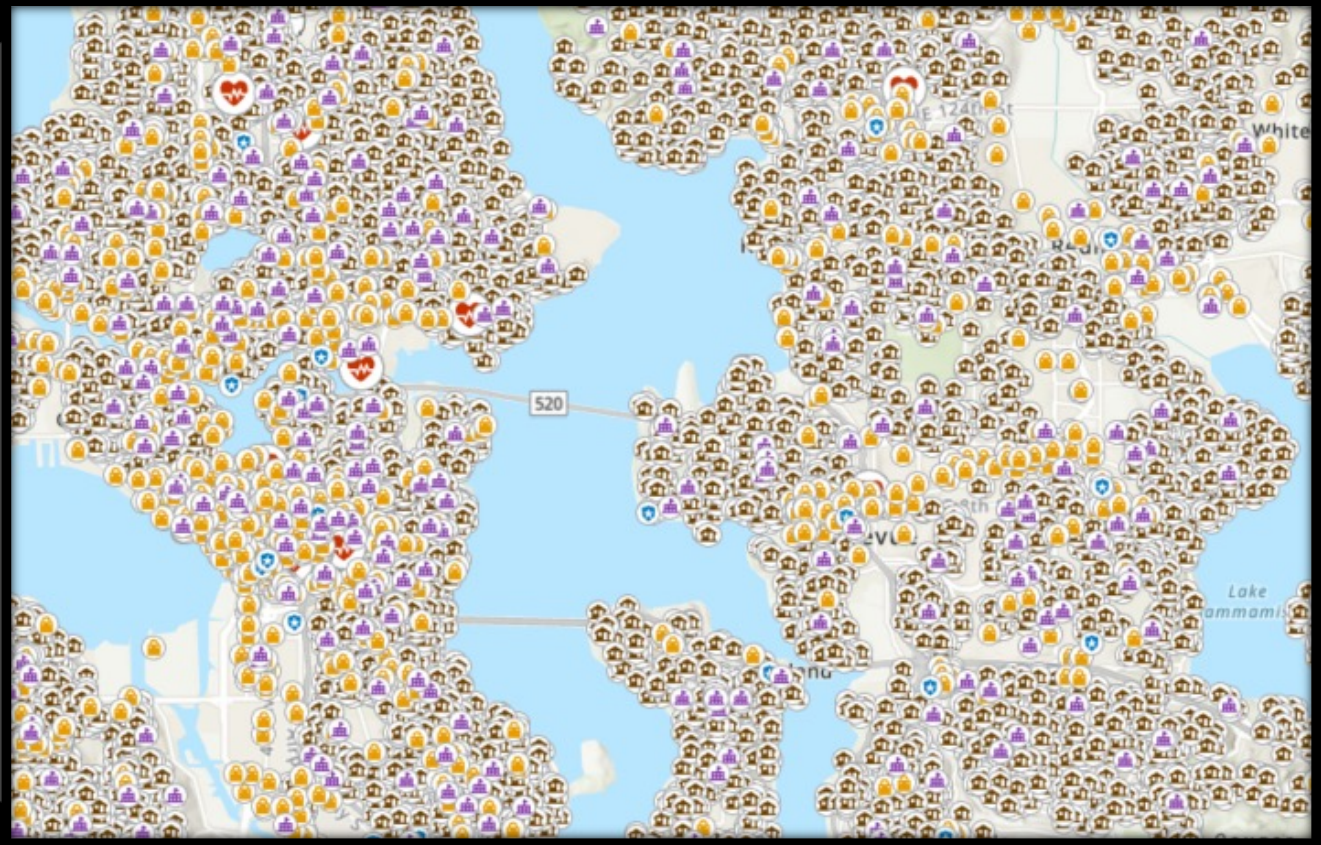
In King County Washington State, USA

- Police Station
- Hospitals
- Food Facilities
- Schools

Ref: King County GIS Open Data

ArcGIS Pro



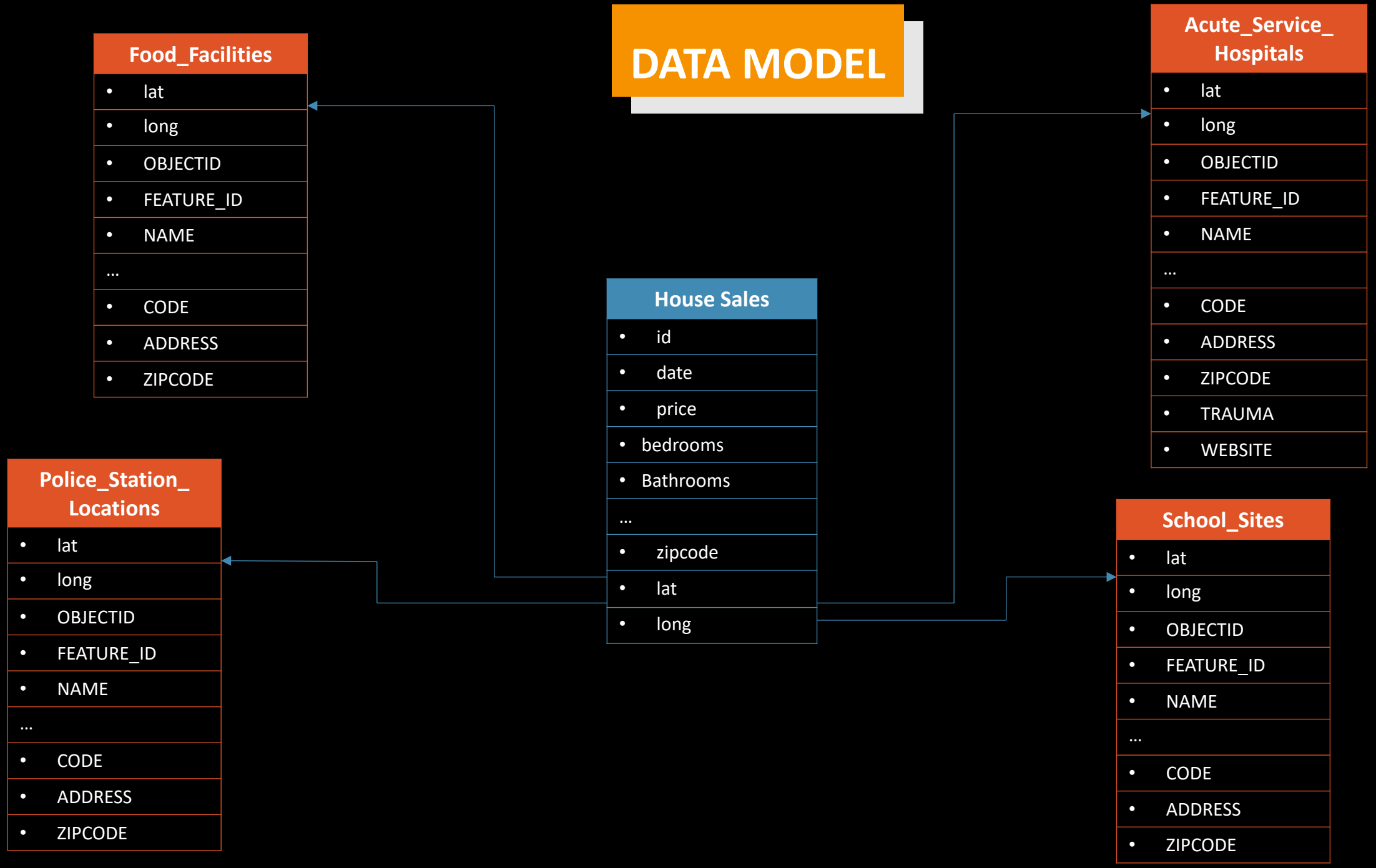


House
School
Shopping Center
Police Station
Hospital

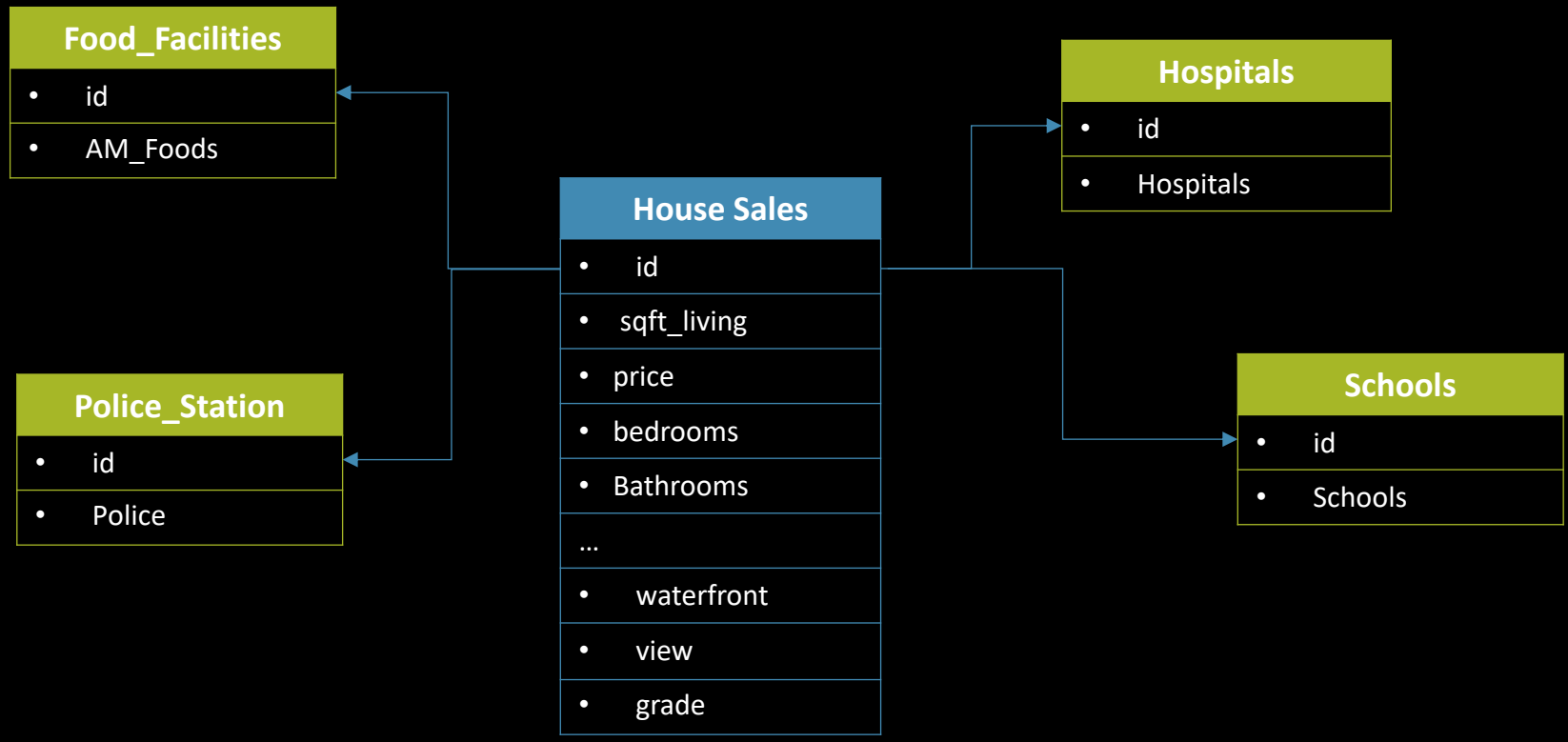

GET DATA
PREPARATION DATA
EXPLORATORY DATA ANALYSIS (EDA)
DATA MODELING
EVALUATION

GET DATA

DATA MODEL


Food_Facilities
lat
long
OBJECTID
FEATURE_ID
NAME
…
CODE
ADDRESS
ZIPCODE
Acute_Service_Hospitals
lat
long
OBJECTID
FEATURE_ID
NAME
…
CODE
ADDRESS
ZIPCODE
TRAUMA
WEBSITE
House Sales
id
date
price
bedrooms
Bathrooms
…
zipcode
lat
long
Police_Station_Locations
lat
long
OBJECTID
FEATURE_ID
NAME
…
CODE
ADDRESS
ZIPCODE
School_Sites
lat
long
OBJECTID
FEATURE_ID
NAME
…
CODE
ADDRESS
ZIPCODE

DATA


Food_Facilities
id
AM_Foods
Police_Station
id
Police
House Sales
id
sqft_living
price
bedrooms
Bathrooms
...
waterfront
view
grade
Hospitals
id
Hospitals
Schools
id
Schools

PREPARATION DATA

DATA TRANSFORMATION

- Determined distance between house and location of alternative data.
- Count location that distance is lower than 5 km.

```
“def distance_from(loc1,loc2):
    dist=hs.haversine(loc1,loc2)
    if round(dist,2) > 5 :
        dist = 0
    else:
        dist = 1
    return dist”
```

```
for _,row in pol_loc.iterrows():
    kc_loc[row.NAME]=kc_loc['coor'].apply(lambda x:
distance_from(row.coor,x))
```

DATA TRANSFORMATION

id	price	lat	long	coor	WSP District 2 Seattle North Detachment	White Center Storefront	KC Airport Law Enforcement	Fairwood Storefront	KC Precinct 3 Headquarters	...	Pacific Police Department	Burien City Hall North Storefront
7129300520	221900.0	47.5112	-122.257	(47.5112, -122.257)	0	0	1	0	0	...	0	1
6414100192	538000.0	47.7210	-122.319	(47.721, -122.319)	0	0	0	0	0	...	0	0
5631500400	180000.0	47.7379	-122.233	(47.7379, -122.233)	0	0	0	0	0	...	0	0
2487200875	604000.0	47.5208	-122.393	(47.5208, -122.393)	0	1	0	0	0	...	0	0
1954400510	510000.0	47.6168	-122.045	(47.6168, -122.045)	0	0	0	0	0	...	0	0

id	pol
7129300520	5
6414100192	2
5631500400	4
2487200875	3
1954400510	2
...	...
263000018	1
6600060120	5
1523300141	5
291310100	1
1523300157	5

MERGE TABLE

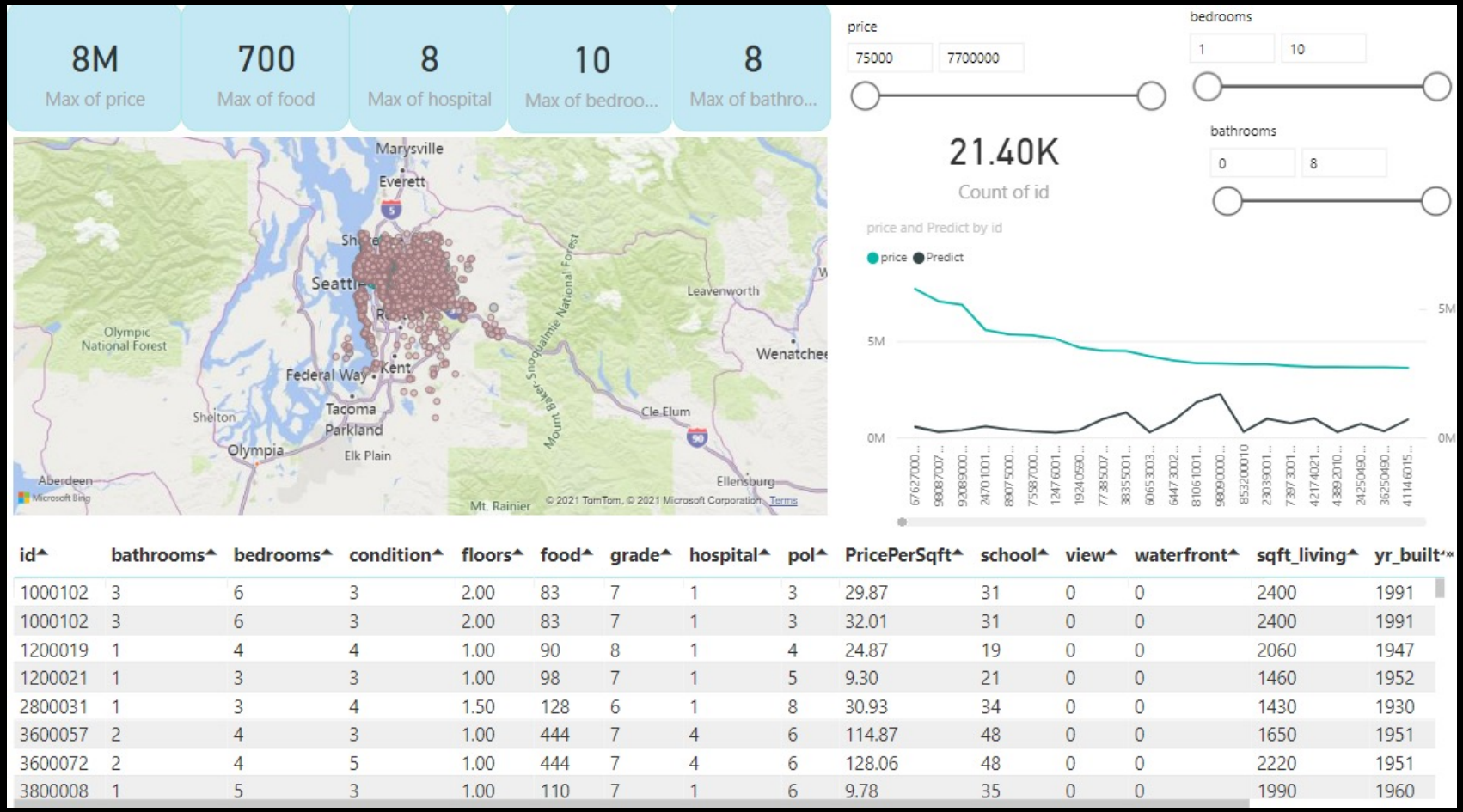
ID	price	bedroom	bathrooms	...	food	school
357668	250000	1	1	...	9	0
345466	350000	2	2	...	8	1
354676	300000	1	1	...	6	6
346533	480000	3	2	...	8	2
...	...	...	...	...	...	...
864367	500000	4	3	...	4	1
856363	538000	5	5	...	7	3

EXPLORATORY DATA ANALYSIS (EDA)

EXPLORATORY DATA ANALYSIS

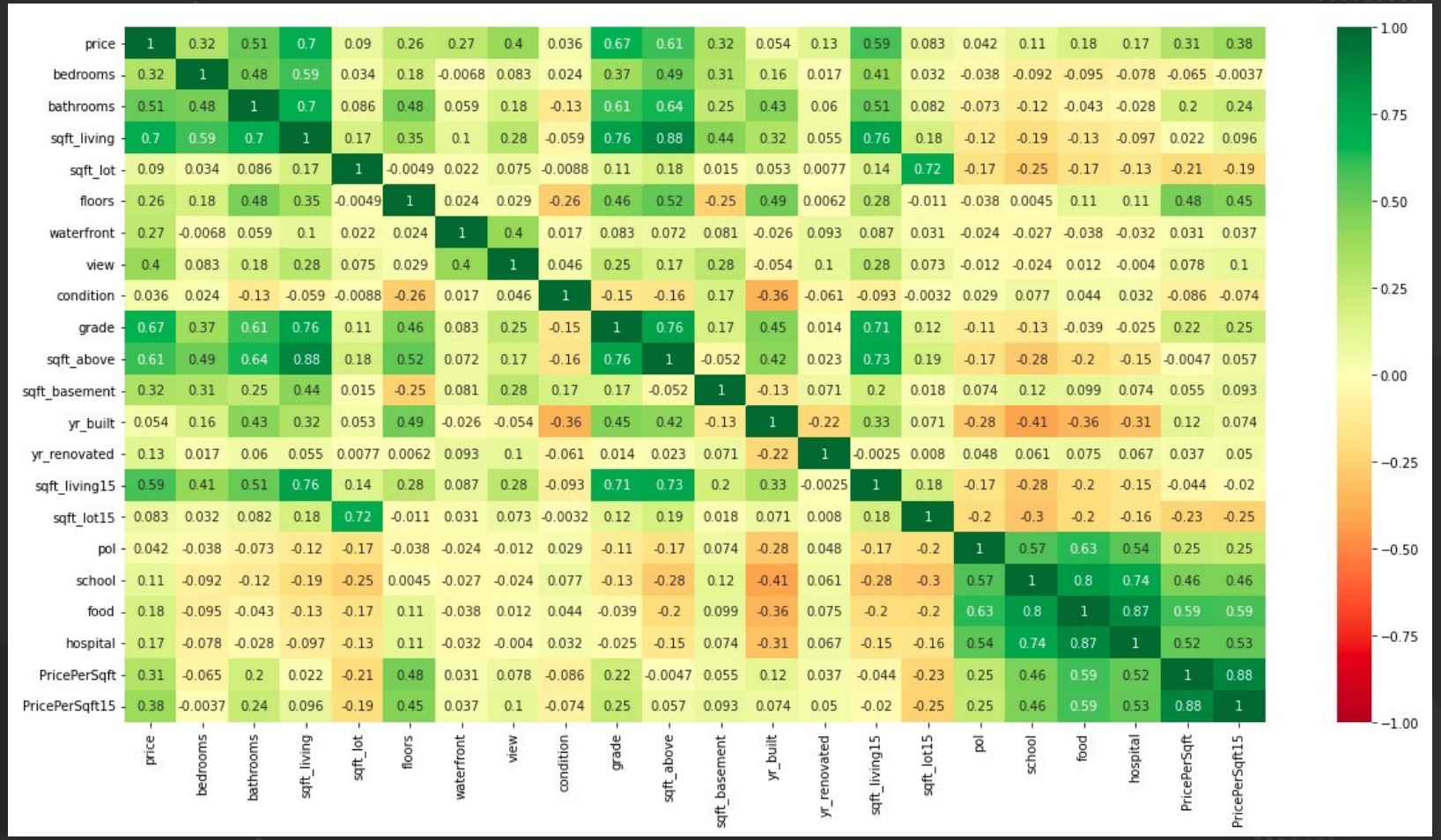
EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS



id	bathrooms	bedrooms	condition	floors	food	grade	hospital	pol	PricePerSqft	school	view	waterfront	sqft_living	yr_built
1000102	3	6	3	2.00	83	7	1	3	29.87	31	0	0	2400	1991
1000102	3	6	3	2.00	83	7	1	3	32.01	31	0	0	2400	1991
1200019	1	4	4	1.00	90	8	1	4	24.87	19	0	0	2060	1947
1200021	1	3	3	1.00	98	7	1	5	9.30	21	0	0	1460	1952
2800031	1	3	4	1.50	128	6	1	8	30.93	34	0	0	1430	1930
3600057	2	4	3	1.00	444	7	4	6	114.87	48	0	0	1650	1951
3600072	2	4	5	1.00	444	7	4	6	128.06	48	0	0	2220	1951
3800008	1	5	3	1.00	110	7	1	6	9.78	35	0	0	1990	1960

EXPLORATORY DATA ANALYSIS



- **SHOW CORRELATION OF EACH ATTRIBUTE**

DATA MODELING

REGRESSION MODELING

- Model 1 - MLP Regressor
- Model 2 - LGB Model
- Model 3 - Random Forest Regression
- Model 4 - Gradient Boosting Regression
- Model 5 - XGB Model
- Model 6 – Decision Tree

EVALUATION

EVALUATION

- ROOT MEAN SQUARED ERROR (RMSE)
- MEAN ABSOLUTE ERROR (MAE)
- R-SQUARED(R^2)

EVALUATION

	Model	RMSE	R2 Score	MAE
0	LGB	50939.189275	0.976227	17071.772647
1	XGB	47729.553943	0.979129	18527.985326
2	Random_Forest	72680.106929	0.951604	23160.928087
3	Decision_Tree	132686.776624	0.838701	41298.544360
4	Gradient_Boost	85537.783602	0.932966	50896.417263
5	MLP	156906.026721	0.774443	100864.691388

CONCLUSIONS
&
THE FUTURE

CONCLUSIONS & THE FUTURE

Model 5 - XGB Model is the best model in this case. Top 5 features impact to housing price is ‘bathroom, sqft_living, grade, sqft_above and sqft_living15’ in main data. From alternative data the police station no impact to housing price

In the future, Finding more alternative data for develop model and performance. Finally Create platform for Finding Buy/Rent House and Condo

Q & A