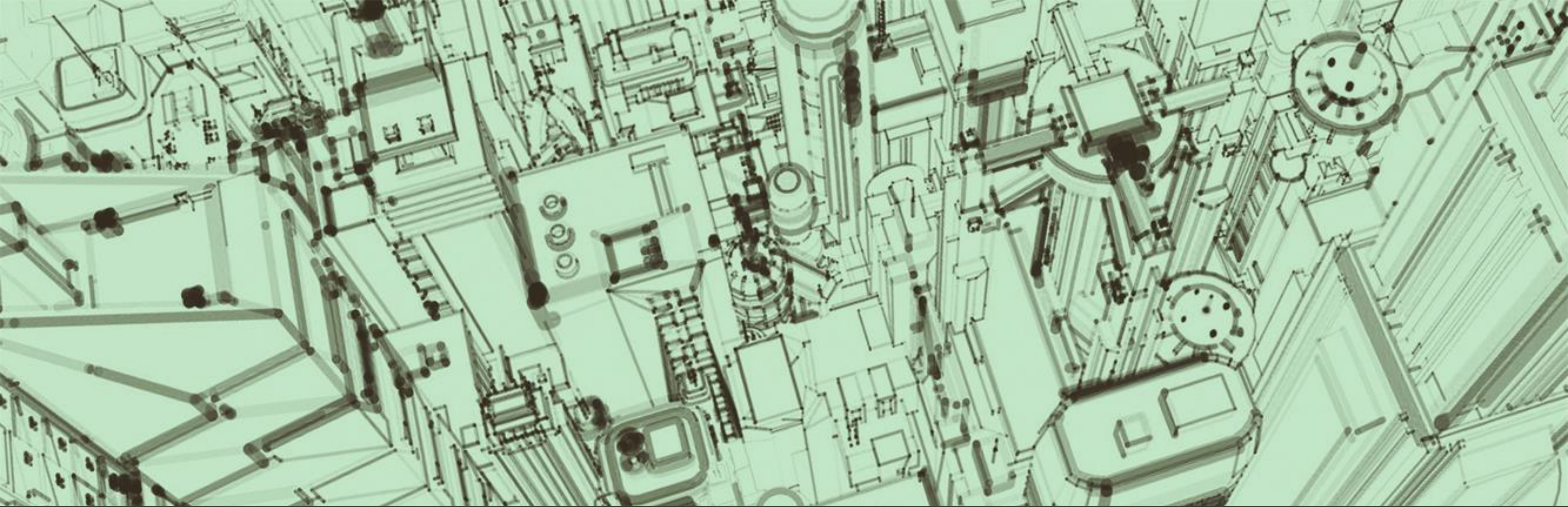


Predicting HOME SALE PRICES in Ames, Iowa using Data Science

Presented by: Daniel Wang

Presented to: Nigerian Prince Shehu Abubakar

Presentation Outline

- Problem Statement
- Discussion of Data & Background
- Primary Findings
- Conclusions
- References
- Questions/Comments?

Problem Statement

- A wealthy Nigerian prince who contacted us online is interested in investing in homes in Ames, Iowa. For a small buy-in fee that I wired to him last week, I've joined his investment group and now am in charge of the project.

- Prince Shehu Abubakar is interested in flipping the homes for profit and is hoping to better understand the factors at play in affecting home sale prices, so that he does not overpay for his investments and can determine fair market values for the properties he wishes to sell.

Data & Background

Our study used data (via [Kaggle]) from the Ames, Iowa Assessor's Office

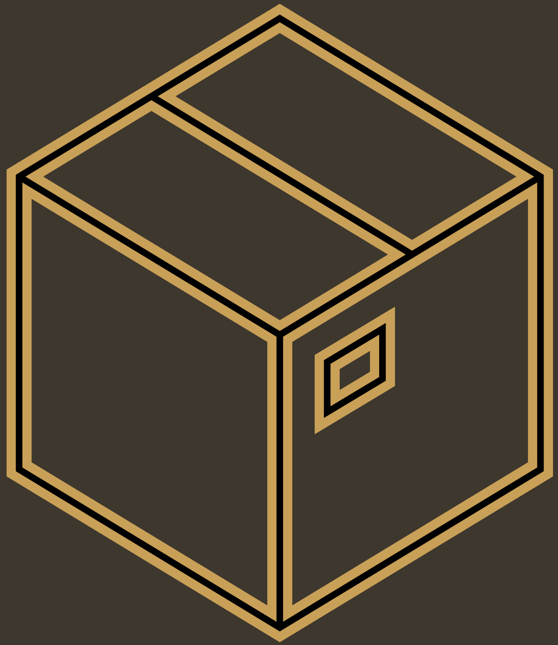
> "used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010",

which contains housing prices and house/property characteristics. These data points are studied and used to create a model that can be used to predict the prices of other homes in the area not included in the original data set.

Data & Background

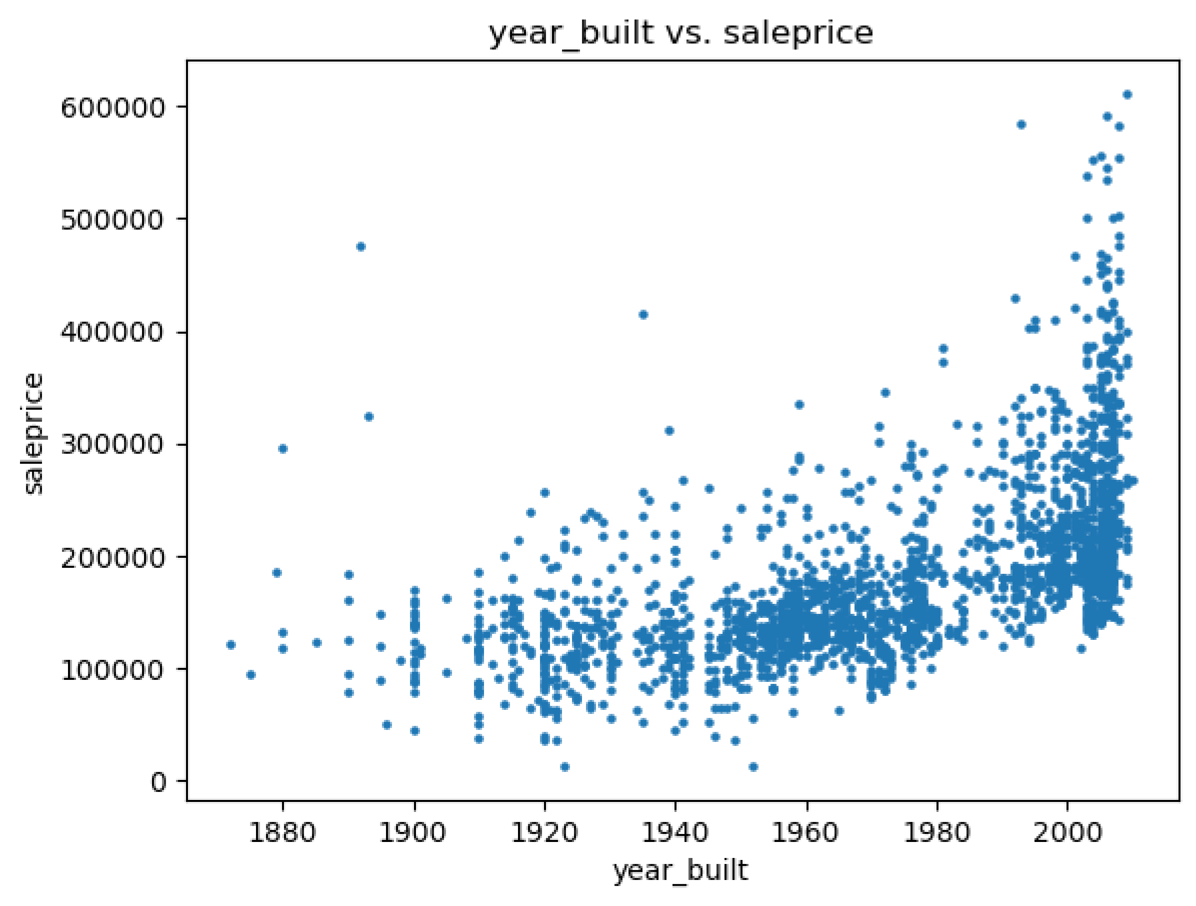
Past Data:

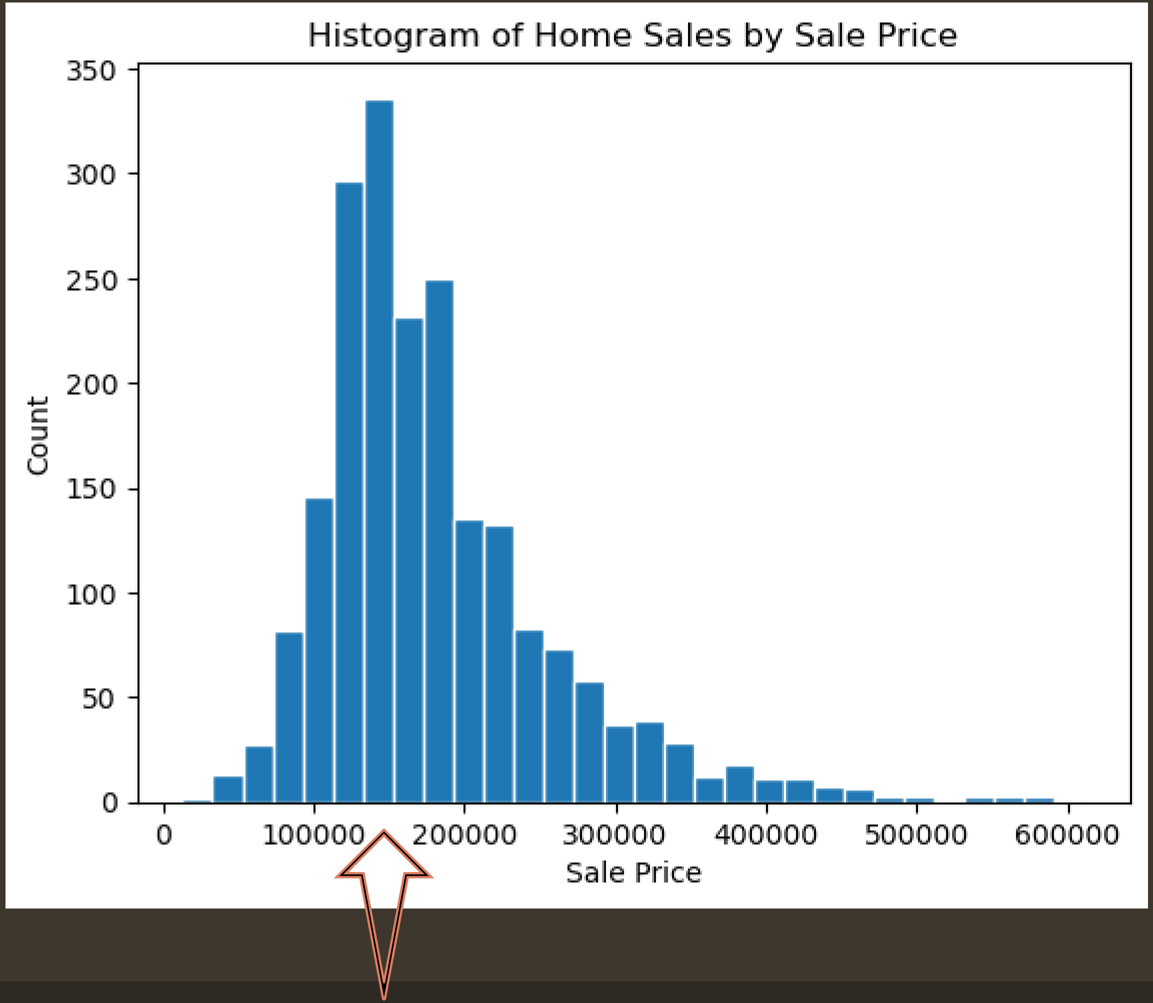
- Overall Quality
- Kitchen Quality
- Heating
- Number of Garages
- Garage Area
- Neighborhood



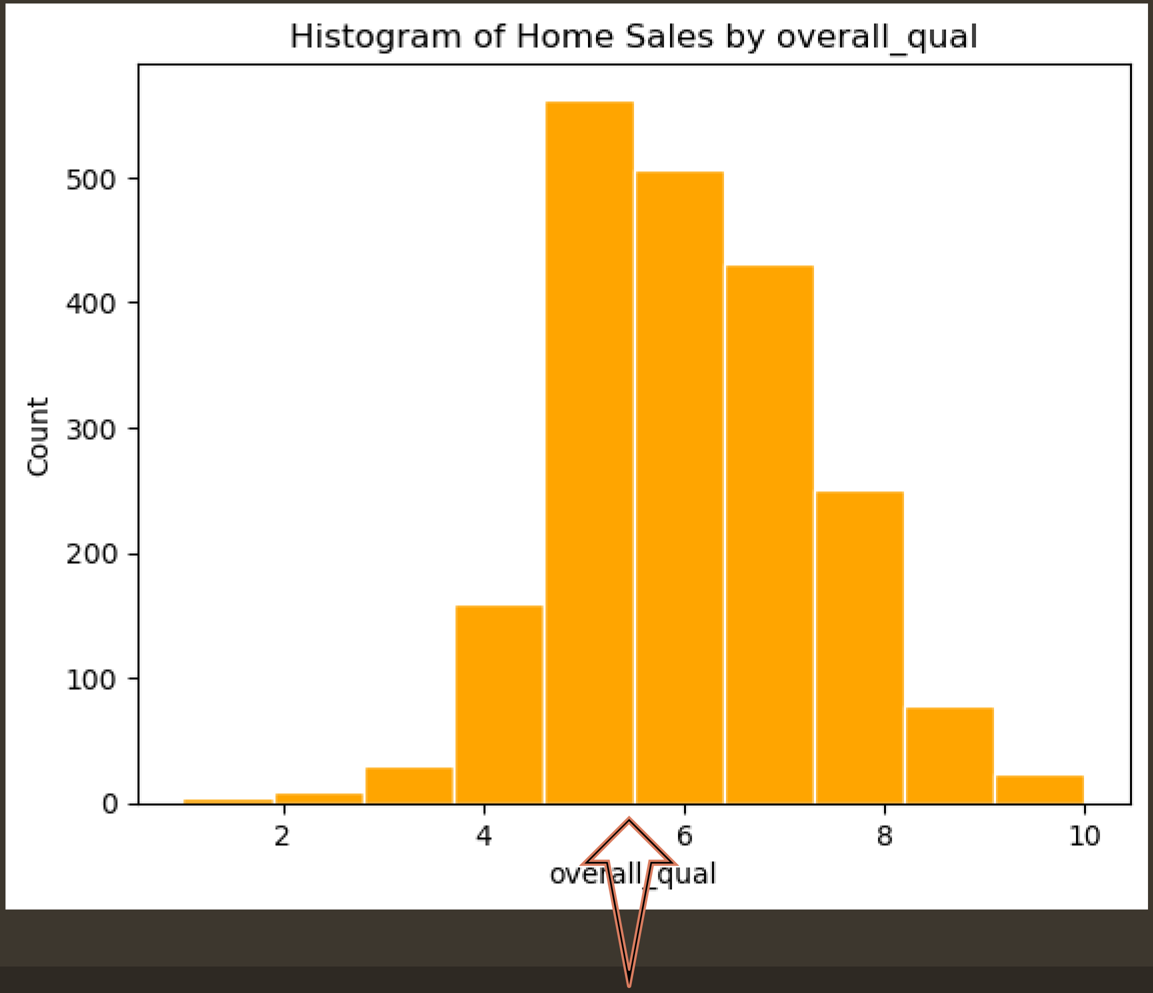
Home Sale Price Predictions!

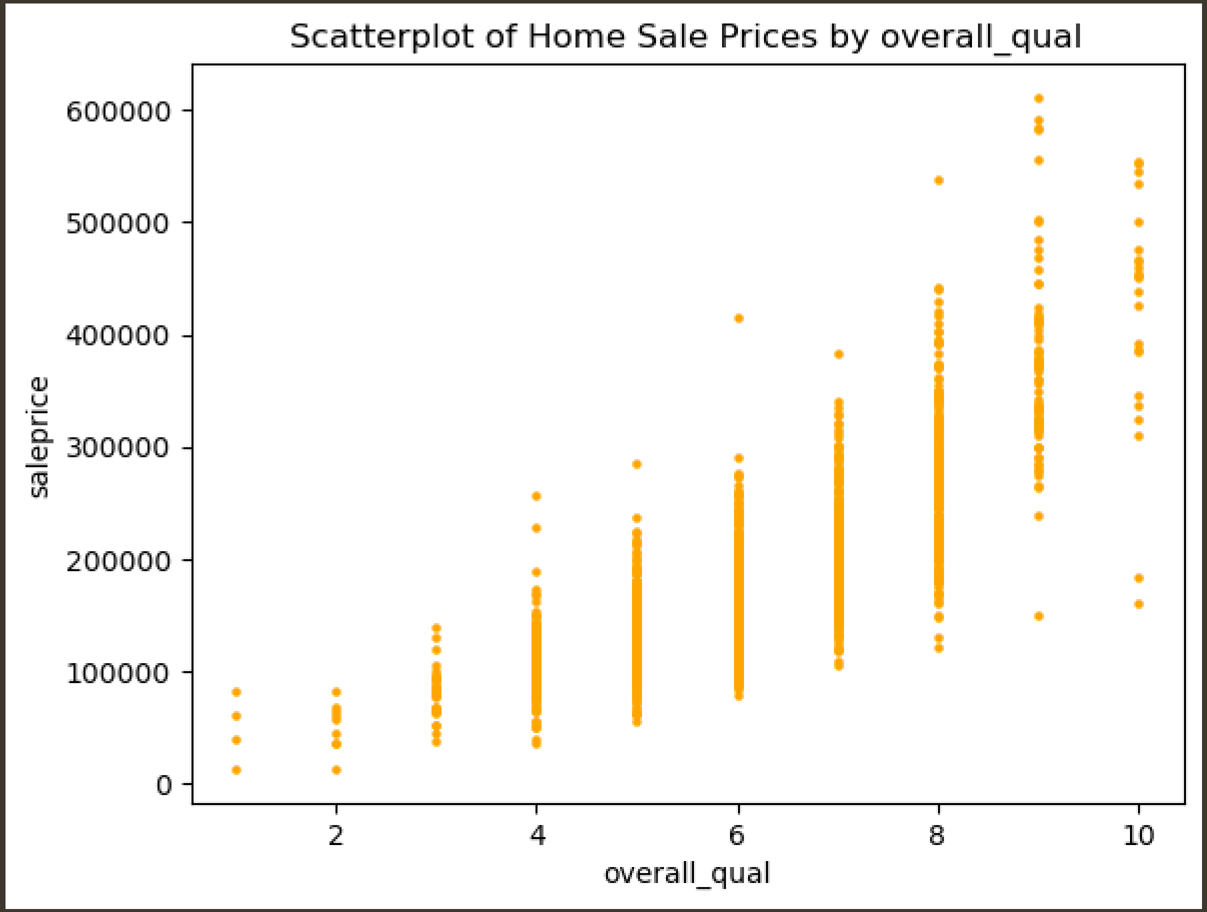
Primary Findings



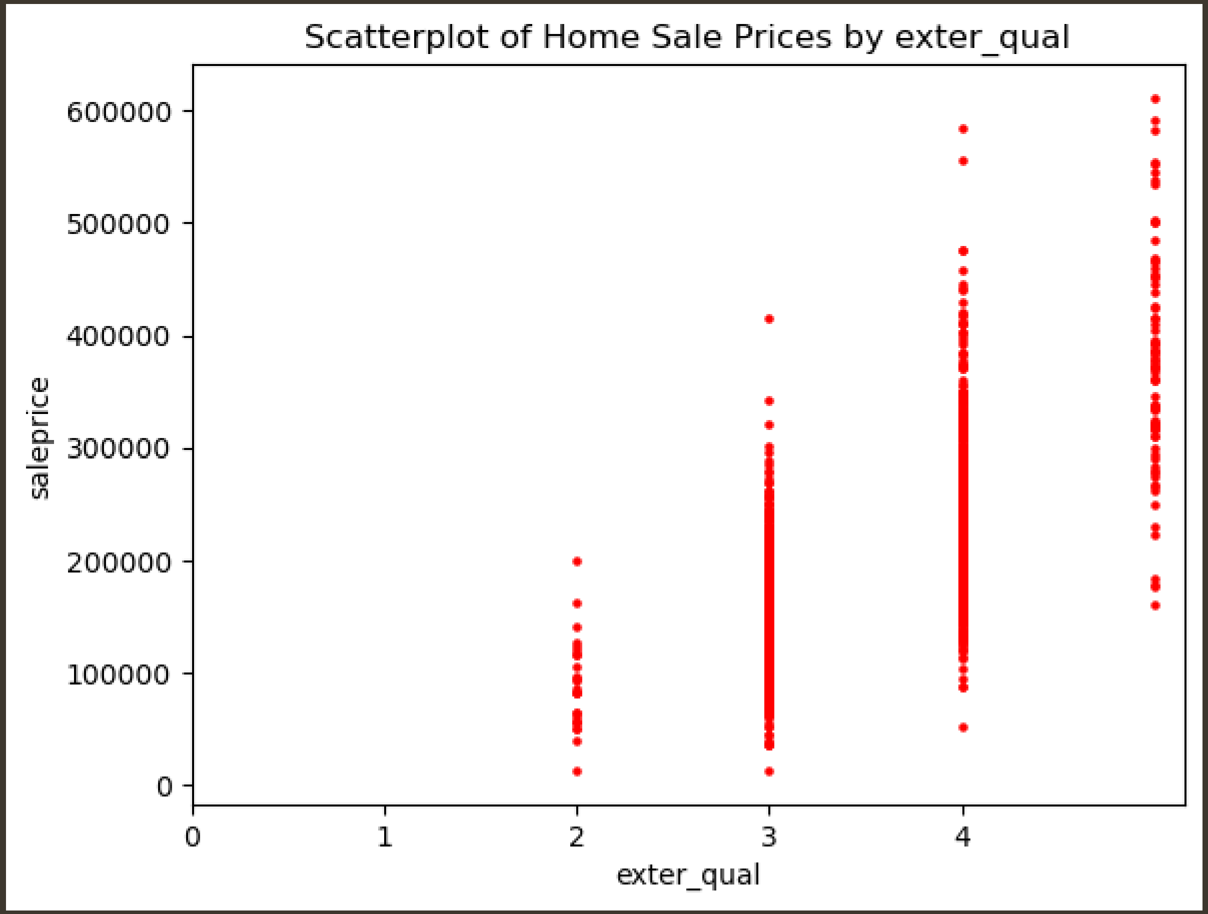


Primary Findings – overall quality

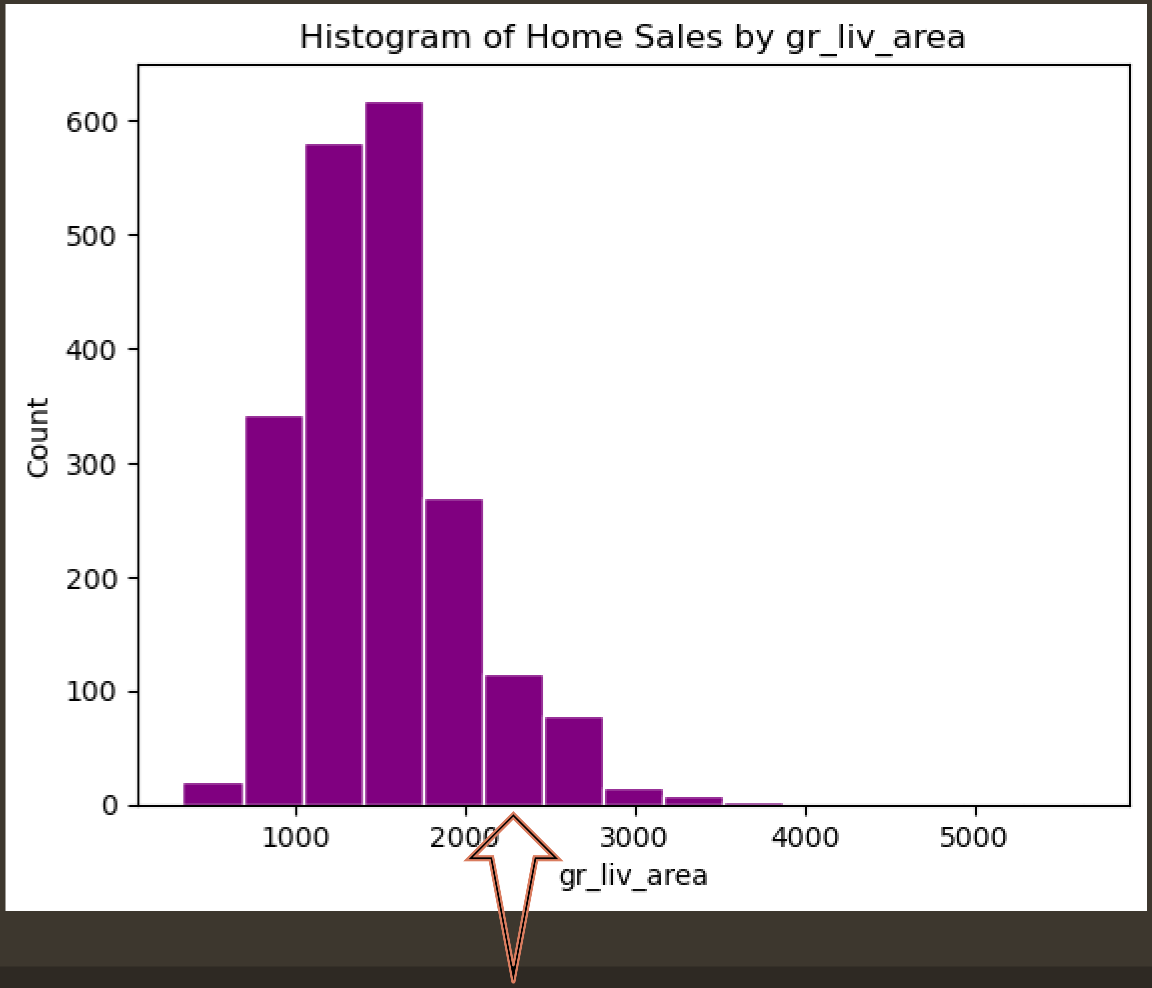


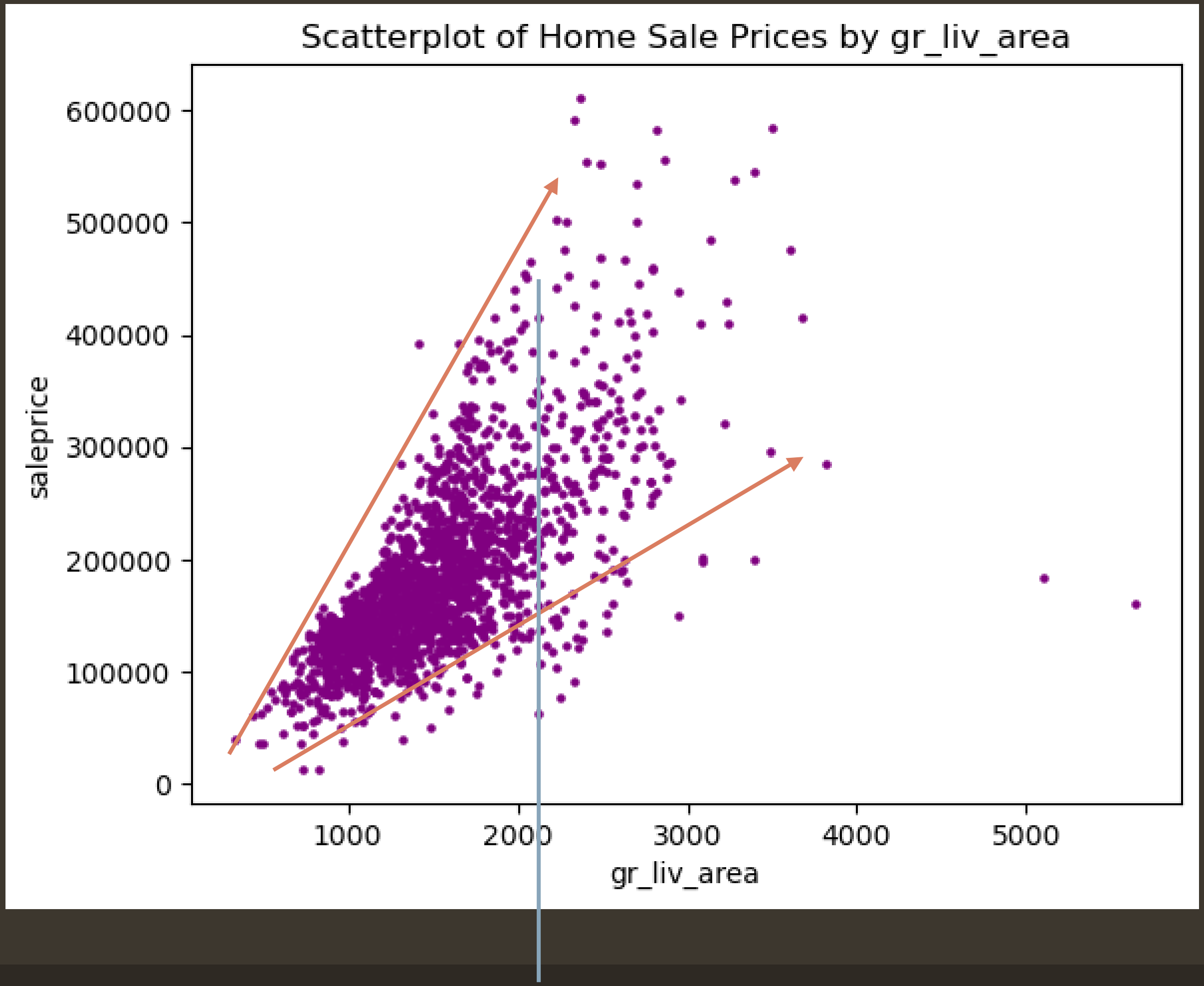


Primary Findings – exterior quality

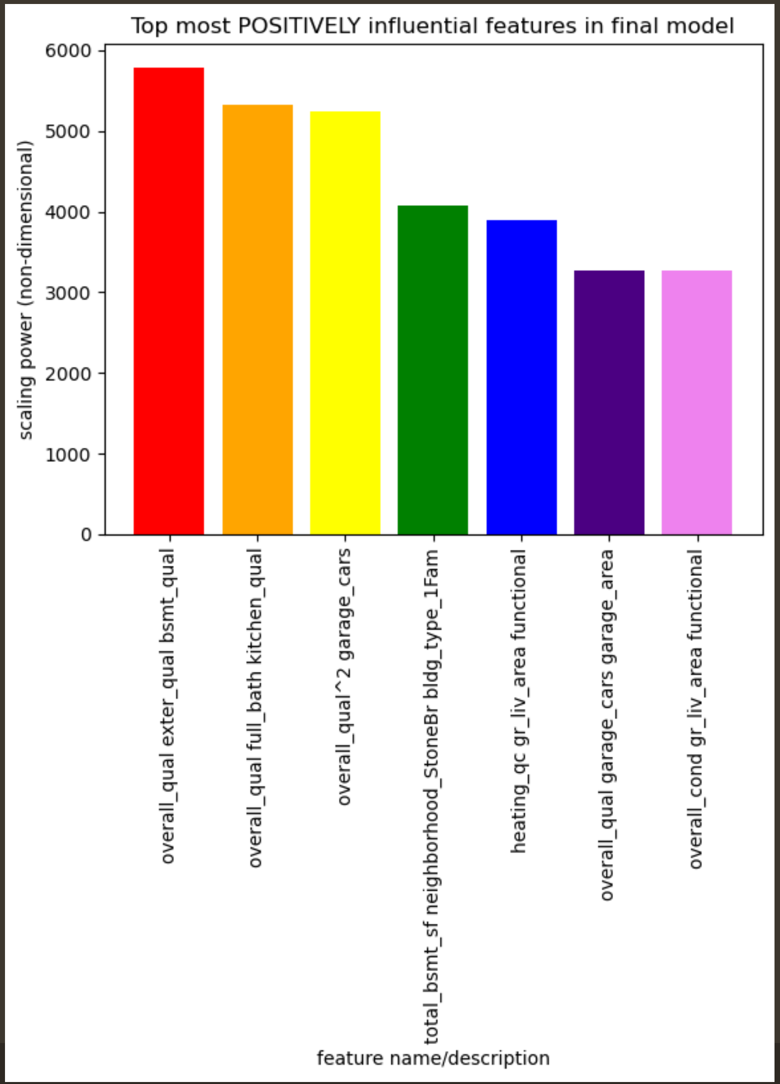


Primary Findings – Above grade (ground) living area square feet



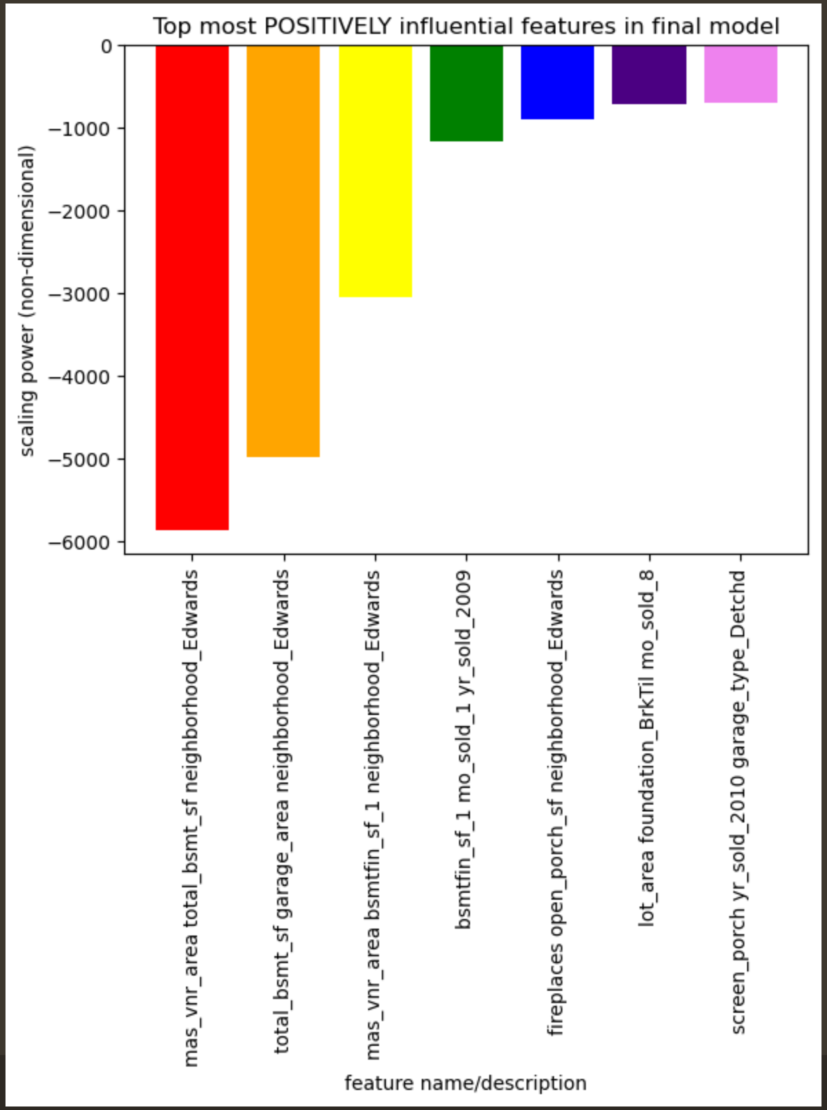


Top 7 most POSITIVELY influential features in model



#	Feature Description	scaling power (non-dimensional)
1	(overall_qual exter) * (qual bsmt_qual)	**5786.106282**
2	(overall_qual) * (full_bath kitchen_qual)	**5328.531647**
3	(overall_qual)^2 * (garage_cars)	**5243.503826**
4	(total_bsmt_sf) * (neighborhood_StoneBr) * (bldg_type_1Fam)	**4078.438986**
5	(heating_qc) * (gr_liv_area functional)	**3898.87032**
6	(overall_qual) * (garage_cars) * (garage_area)	**3275.160347**
7	(overall_cond) * (gr_liv_area functional)	**3265.027374**

Top 7 most NEGATIVELY influential features in model



#	Feature Description	scaling power (non-dimensional)
-1	(mas_vnr_area) * (total_bsmt_sf) * (neighborhood_Edwards)	-5858.773851
-2	(total_bsmt_sf) * (garage_area) * (neighborhood_Edwards)	-4979.10532
-3	(mas_vnr_area) * (bsmtfin_sf_1) * (neighborhood_Edwards)	-3049.396129
-4	(bsmtfin_sf_1) * (mo_sold_1) * (yr_sold_2009)	-1170.813491
-5	(fireplaces) * (open_porch_sf) * (neighborhood_Edwards)	-899.948678
-6	(lot_area) * (foundation_BrkTil) * (mo_sold_8)	-716.521048
-7	(screen_porch) * (yr_sold_2010) * (garage_type_Detchd)	-709.927402

Conclusions

- Housing prices can be modeled and predicted using data science techniques and past data.

- Top (individual) factors correlated with saleprice appear to be:

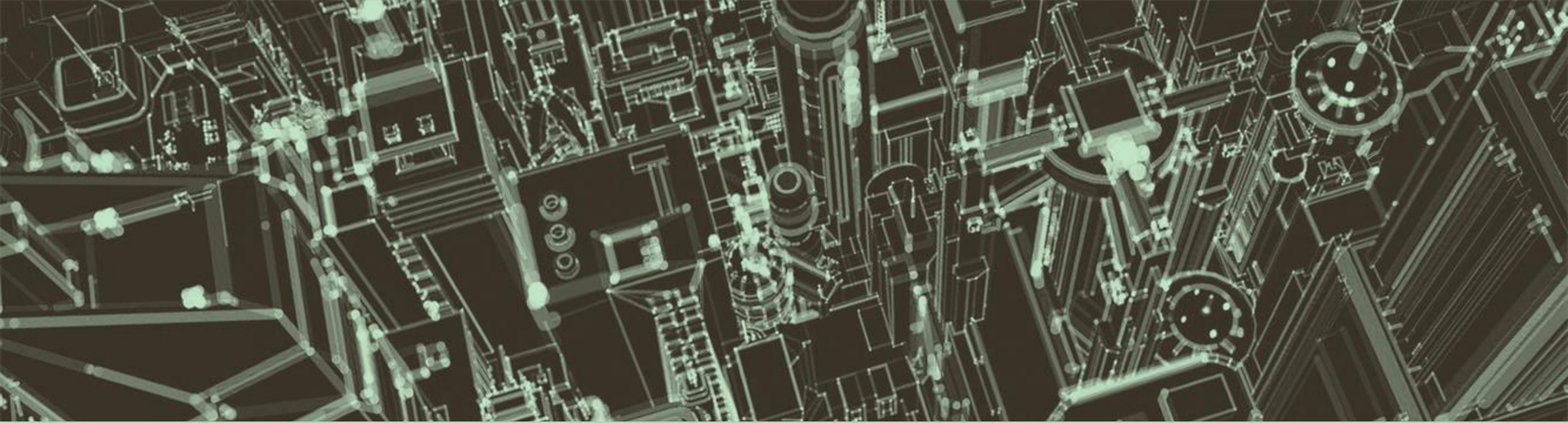
  overall_qual, exter_qual, gr_liv_area, kitchen_qual, garage_area.

  It stands to reason that for a person hoping to increase their home sale value, these would be the areas of focus.

- Top higher-order (3rd degree polynomial) factors correlated with saleprice are listed in the previous tables

References

1. Kaggle
2. Ames, Iowa Assessor's Office – original source for data dictionary information
   - http://jse.amstat.org/v19n3/decock/DataDocumentation.txt



Questions/Comments?

Thank you for your time!