

# Predicting House Prices in Ames, Iowa

Elizabeth Ayisi, Lauren Tipple, & Dursun Caliskan

Data Analytics Final Project  
Summer 2022

# Rationale

1. The data retrieved from kaggle was clean to work with for our intended project.
2. The data retrieved from kaggle seemed manageable for principal component analysis.



# Description of Data

- 81 columns of homes features.
- 1460 rows.
- Half of the data was categorical and the other half, numerical.



# Research Questions

- 01 To what extent does the machine learning model predict house prices in Iowa.
- 02 How accurate are the different linear regression models?



# Description of the Data Exploration Phase

1. Database - PostgreSQL was used for the database. We used Google Colab, pandas and pySpark to explore the data, extract, transform, and load into the database. The database was created using AWS, RDS.
2. We used pgAdmin to create the table schema in RDS.
3. Used Spark on Colab to clean and transform the data.
4. Loaded the data from Pandas DataFrames into RDS.



# Description of the Analysis Phase

- Linear Regression
- Chi-Square test
- Simple imputation & KNN method (this was the most accurate after comparing the two methods)
- Principal Component Analysis

# Technologies, Languages, Tools, and Algorithms Used

AWS

S3

Spark

Pandas

RDS

Scipy

Sklearn.impute

Python

Seaborn

Pyspark

Google collab

PostgreSQL

Numpy

Scipy.stats

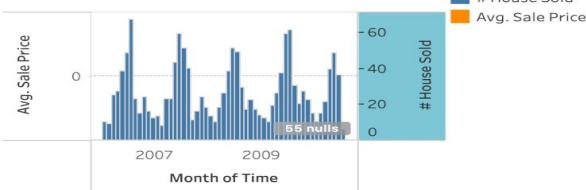
Google Slides



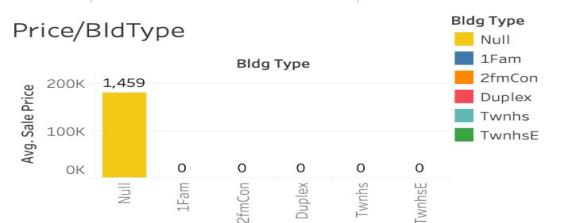
# Storyboard

## Predicting House Prices in Ames, Iowa

Time Series



Price/BldType



GarageArea/Cars/SalePrice



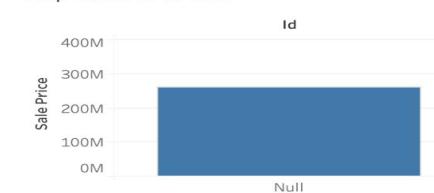
Sales Prices Distribution



GrLivingArea



ExpensiveHouse



Time Series

Sales Prices Distribution

Price/BldType

GrLivingArea

GarageArea/Cars/SalePrice

ExpensiveHouse

CheapHouse

Pool

Dashboard 1

Story 1



# Results of Analysis

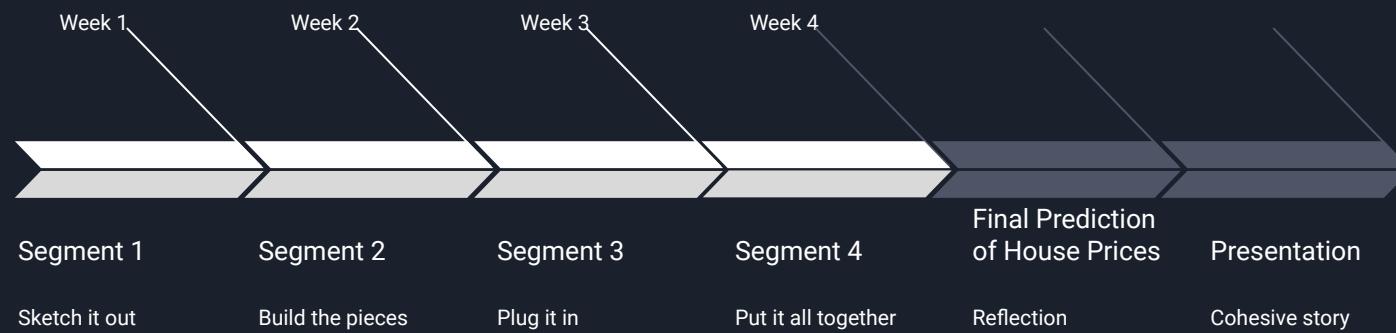


# Recommendation for Future Analysis



# Anything the Team Would Have Done Differently

# Project timeline





Description of the tool(s) that will be used to create the final dashboard



# Description of interactive element(s)



## References

Dean De Cock (2022). House prices - Advanced Regression techniques. *Kaggle*. Retrieved from <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

Thank you!

