FLIP ROBO

House Prediction Model

Submitted by:

Siva Kumar V

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

In house prediction model number feature is more so we need to understand about many technique.

Geek4Geeks and towards data science and google is given lot of technique to resolve the issue.

INTRODUCTION

- ## Business Problem Framing

You are required to model the price of houses with the available independent variables. This model will then be used by the management to understand how exactly the prices vary with the variables. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns. Further, the model will be a good way for the management to understand the pricing dynamics of a new market.

- ## Conceptual Background of the Domain Problem

We need to have house building and construction related domain knowledge to deal this project.

- ## Review of Literature

We need to handle

- Null values
- Skewness of the data
- Removing outlier
- Feature dimension reduction
- Find best model

- ## Motivation for the Problem Undertaken
  Describe your objective behind to make this project, this domain and what is the motivation behind.

Houses are one of the necessary need of each and every person around the globe and therefore housing and real estate market is

one of the markets which is one of the major contributors in the world's economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company.

Analytical Problem Framing

- ## Mathematical/ Analytical Modeling of the Problem

  Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

  Size , grouping and finding average of the model.

  Relation between two feature is found.

- ## Data Sources and their formats

  What are the data sources, their origins, their formats and other details that you find necessary? They can be described here. Provide a proper data description. You can also add a snapshot of the data.

  Total data source size of train data is 1168, 81 and test data is 292 ,80 . Data source format is CSV

- ## Data Preprocessing Done

  What were the steps followed for the cleaning of the data? What were the assumptions done and what were the next actions steps over that?

  We are Pre-processed 4 kind of feature discrete, continues, categorical and temporarily variable.

- Data Inputs- Logic- Output Relationships

  Describe the relationship behind the data input, its format, the logic in between and the output. Describe how the input affects the output.

  Our target variable having linear relation with independent variable.

- State the set of assumptions (if any) related to the problem under consideration

  Here, you can describe any presumptions taken by you.

- Hardware and Software Requirements and Tools Used

  Listing down the hardware and software requirements along with the tools, libraries and packages used. Describe all the software tools used along with a detailed description of tasks done with those tools.

  Hardware : Windows /Mac Os , 4Gb Ram,minimum 100 gb rom , minimum Pentium dual core or intel i3 processor.
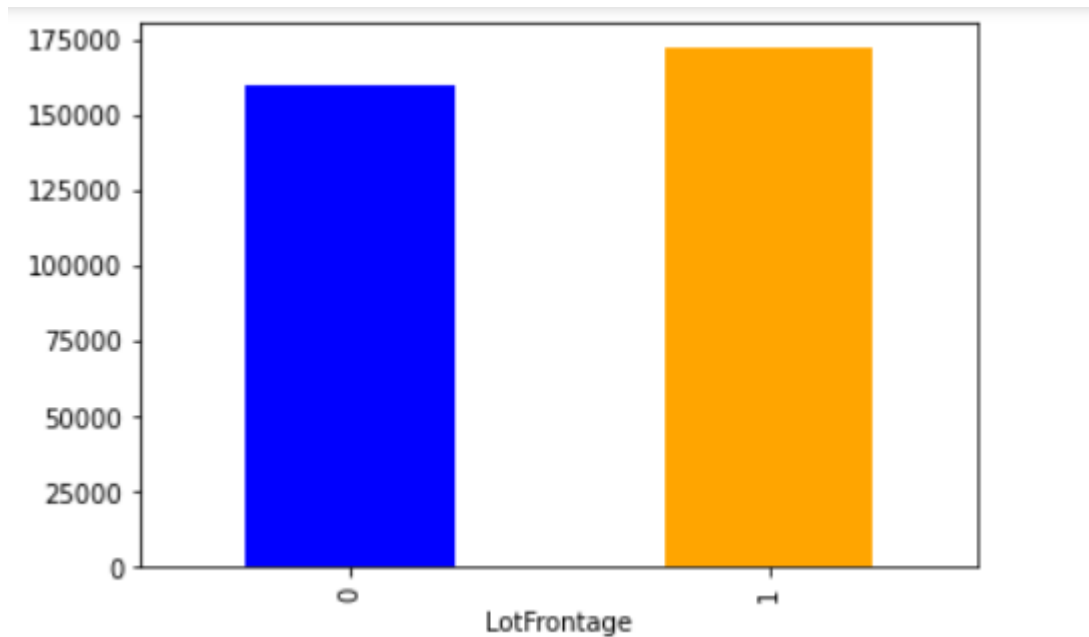
  Software: Jupyter Notebook ,Python and other ML libraries.
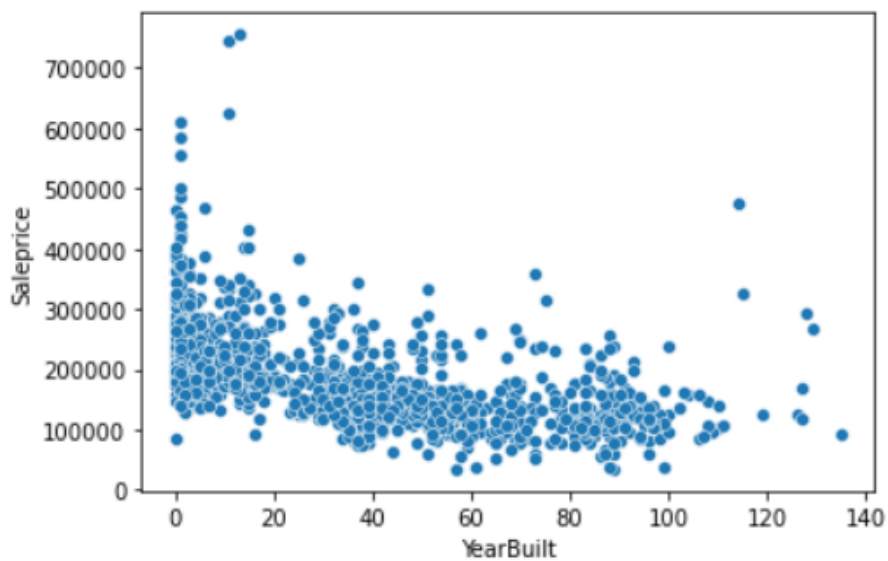
Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

  Describe the approaches you followed, both statistical and analytical, for solving of this problem.

  We identified linear regression model is suitable for train this model.

- Testing of Identified Approaches (Algorithms)

  Listing down all the algorithms used for the training and testing.

  Used linear regression model to test this model.

- Run and Evaluate selected models

  Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

  Checked performance using R2 score.

- Key Metrics for success in solving problem under consideration

  What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

  Checked performance using R2 score.

- Visualizations

  Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

  If different platforms were used, mention that as well.

  To check the null value.

Checking relation with NA and Sales Price

Relationship between Year feature and Sales Price

- ## Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

Using the above visualizations, preprocessing and modelling we found relationship between data and clean the feature , train the data using suitable model for house price prediction.

CONCLUSION

- ## Key Findings and Conclusions of the Study

  Describe the key findings, inferences, observations from the whole problem.

  Learned how to handle more number of feature in the given dataset.

- ## Learning Outcomes of the Study in respect of Data Science

  List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

  1. Understand about relationship between feature and target
  2. Use of cleaning dataset
  3. Train the model and implement in the real world application.

- ## Limitations of this work and Scope for Future Work

  What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.

  1. We should understand each feature manually
  2. Data pre-processing very time taking process.
  3. We need to handle original dataset carefully or else complete model will fail.