

PROJECT: HACK9



Challenge 11: Cost Prediction Model TEAM 11-B:

Participants: Gianmarco Mendiola, Tong Zhi, Younes El Khatib, Chen-Chi Yun, Valeria Rodriguez, Brandon Chin Woei Lim, Zainab Kapasi



@pdacommunity



Projecting Success



Project Data Analytics
Community



Tackling the Problem



ML model selection

- Identify and implement a suitable machine learning model that fits the historical data and provides accurate predictions.



Interactive cost forecasting

- Use user inputs to predict contract costs of future projects.



Automated Predictive tool

- Visual dashboard integrated with the ML model to produce the user an interactive cost prediction.

Process Flow

01

Exploratory Data Analysis (EDA) to identify insights from data

02

Implementation of **machine learning** (XGBoost) and **clustering** algorithms

03

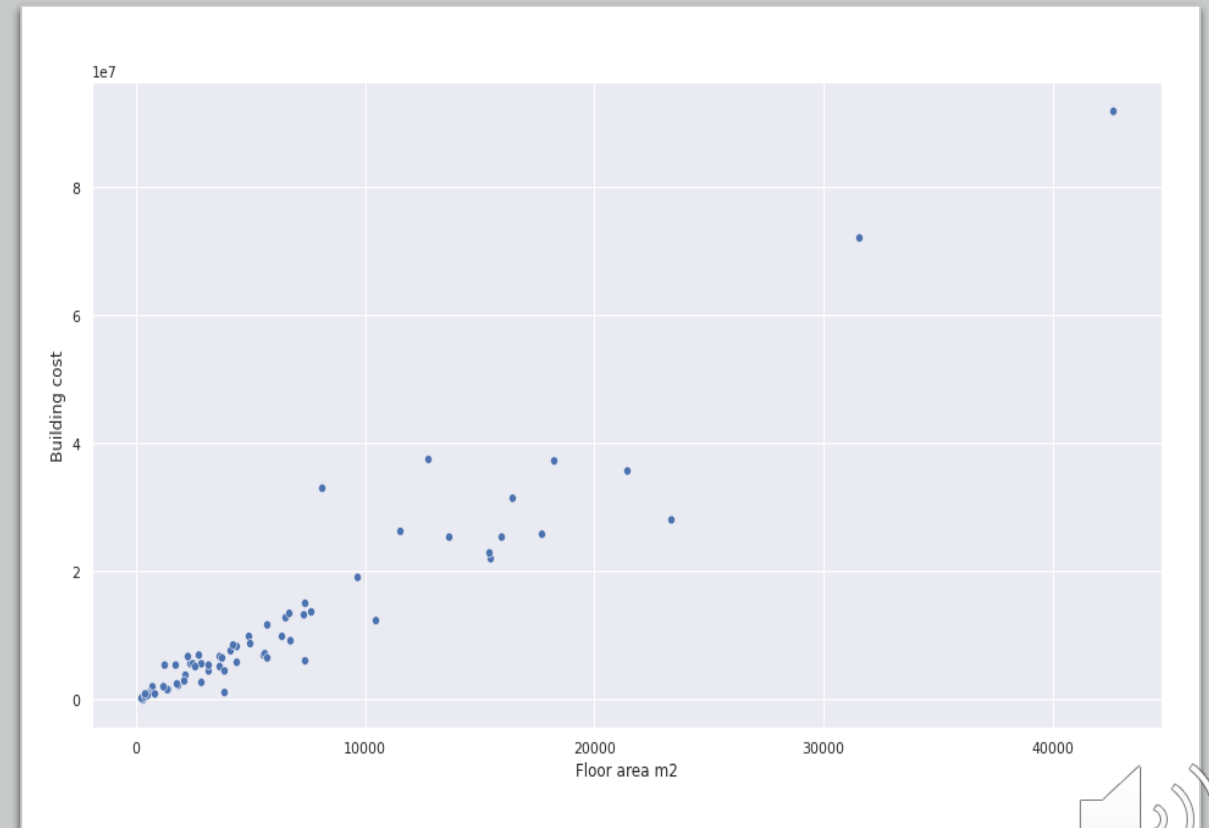
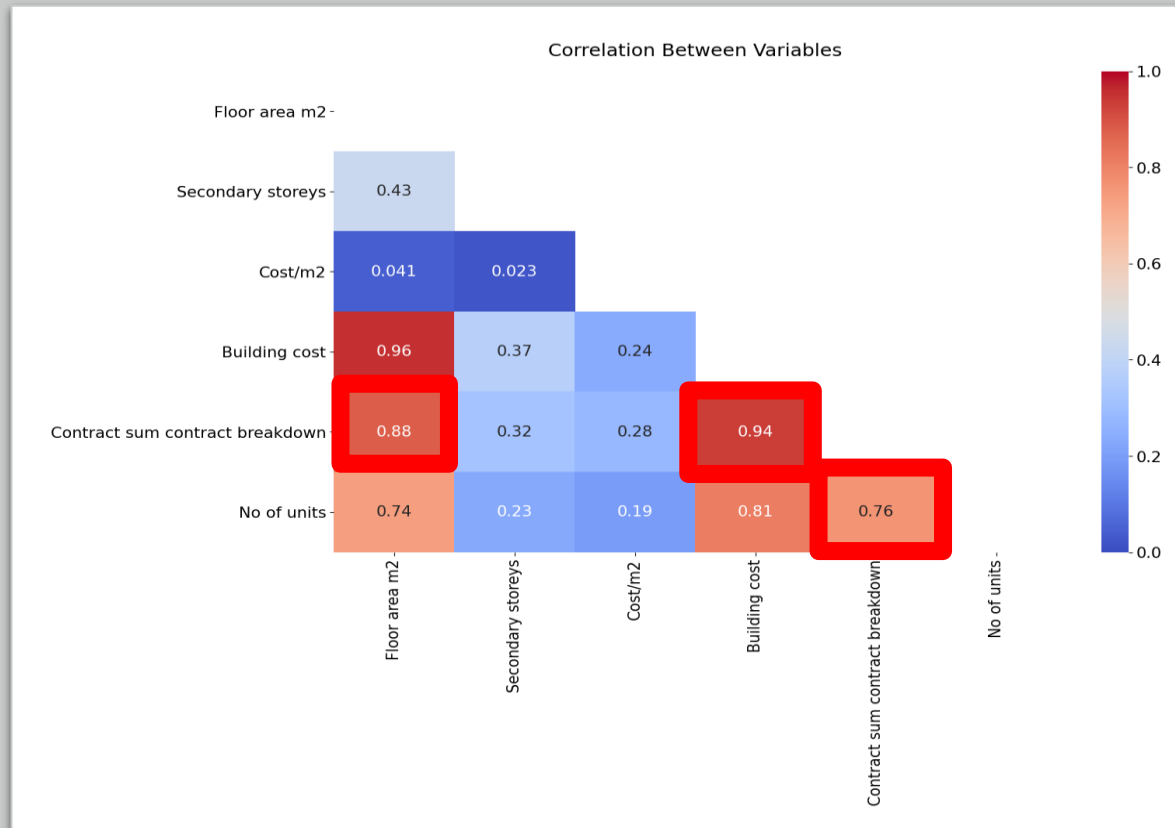
Automated a user friendly **dashboard in Power BI**



Exploratory Data Analysis

Feature Reduction: Numerical Correlation Analysis

- Building cost variable is highly correlated with Cost/FU , floor area.
- Floor Area vs Building cost scatter plot shows positive relationship.

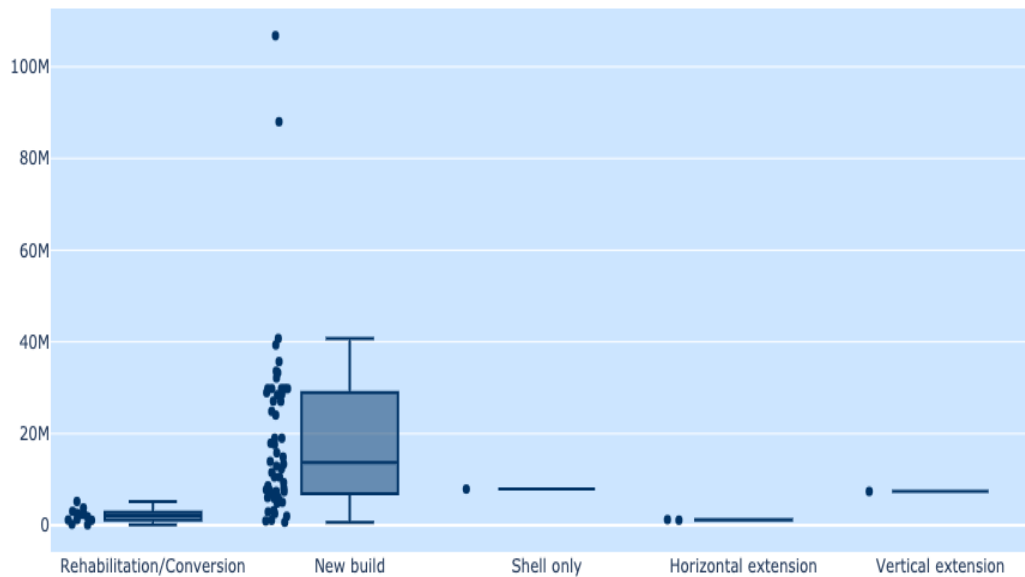


Exploratory Data Analysis

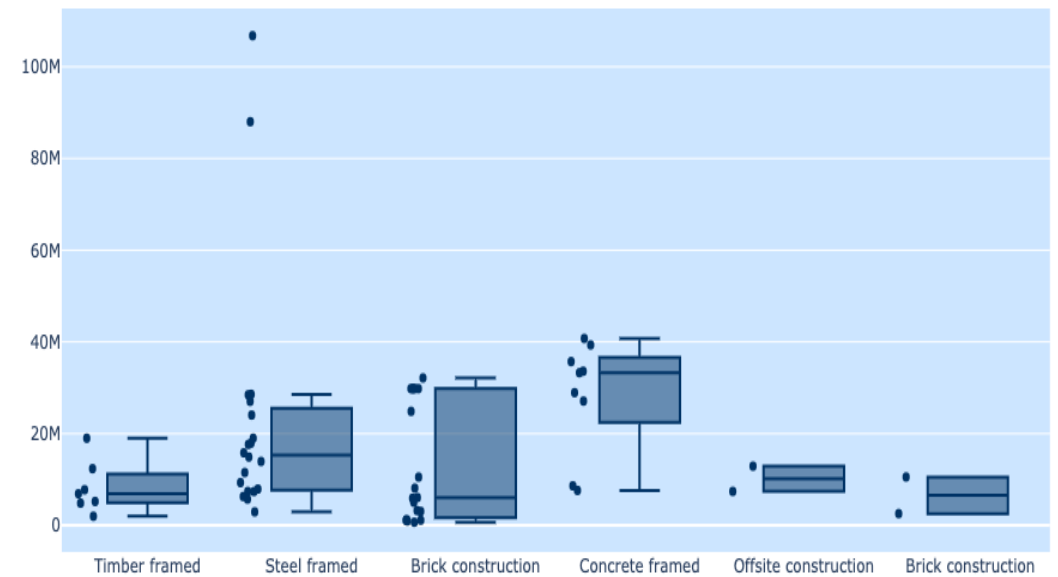
Categorical Variable impact

- Cost is higher when the construction is a new building
- Cost increases when the building is made of steel, brick or concrete

Cost Vs Building Usage



Cost Vs Construction Type



Machine Learning Model

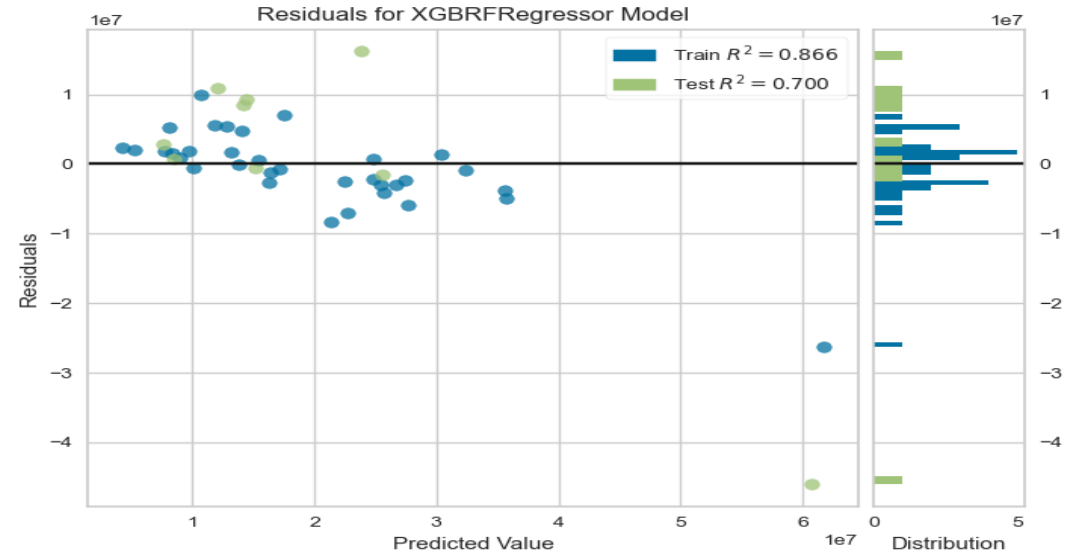
Extreme Gradient Boosting Random Forest (XGBRF)

Diagnostics and Limitations

- Model performance limited by the size and amount of quality training data.
- Subpar model performance for extreme responses in training data.

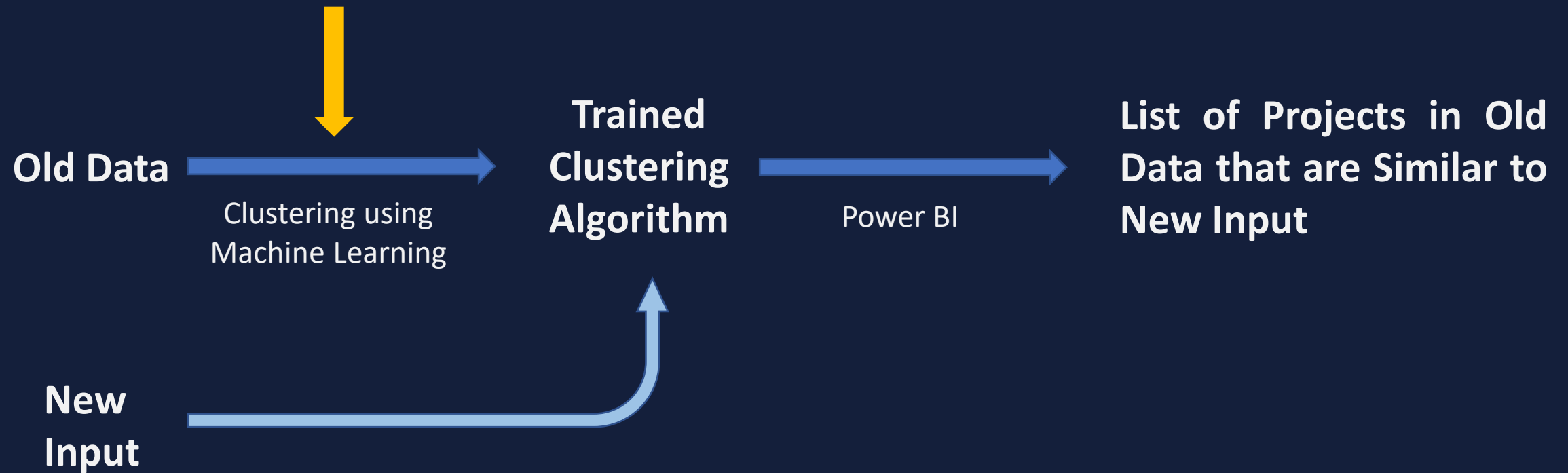
Advantages

- Comparatively robust performance for small datasets.
- Handles “dirty” data and prevents overfitting.
- Fast and easy implementation .



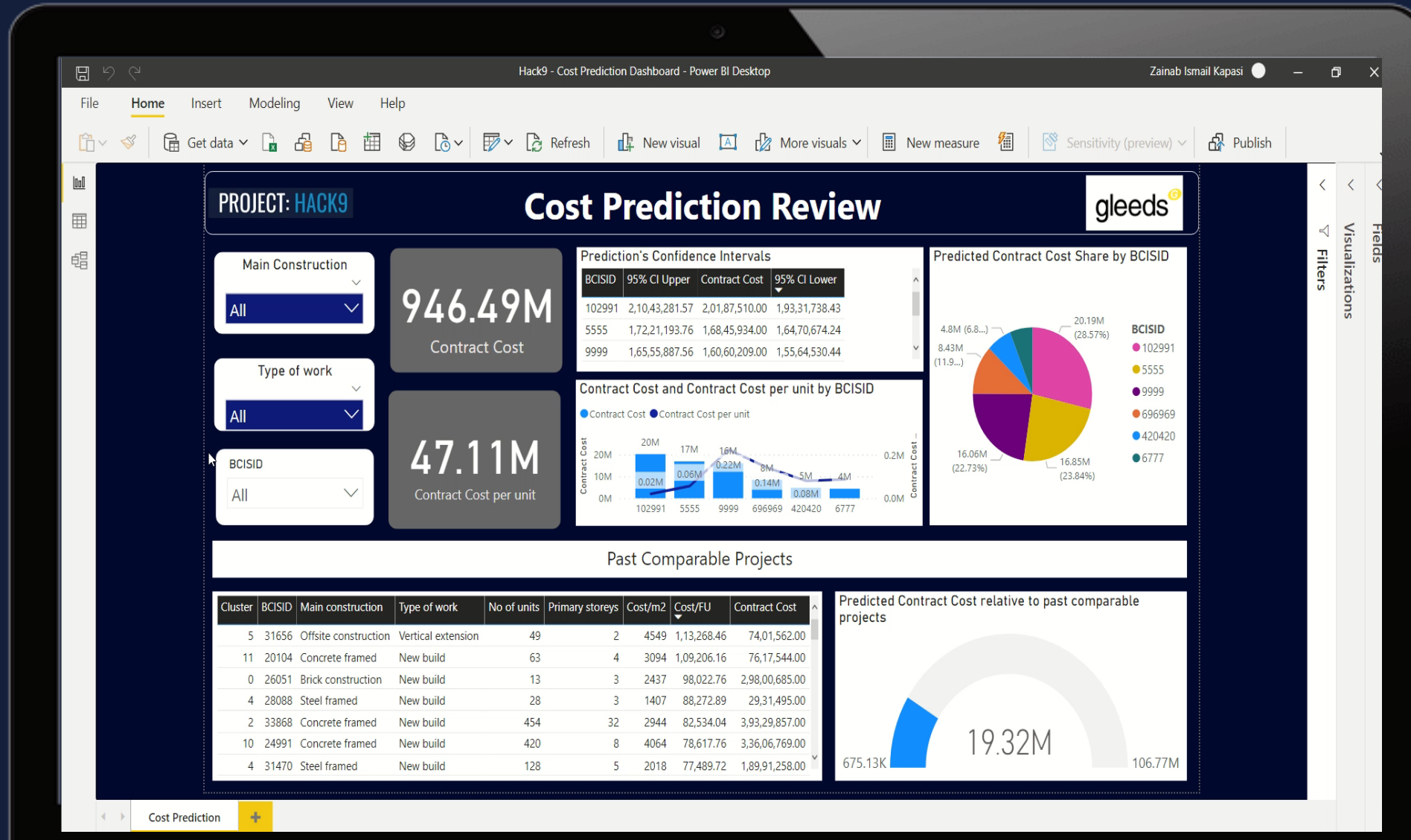
Finding Similarities Between New Inputs and Old Data

Future Consideration for Higher Dimensional Inputs
Integrate with t-Distributed Stochastic Neighbour Embedding



Cost prediction interactive tool

Power BI | Python | Excel





PROJECT DATA ANALYTICS

Questions

PROJECT: HACK9

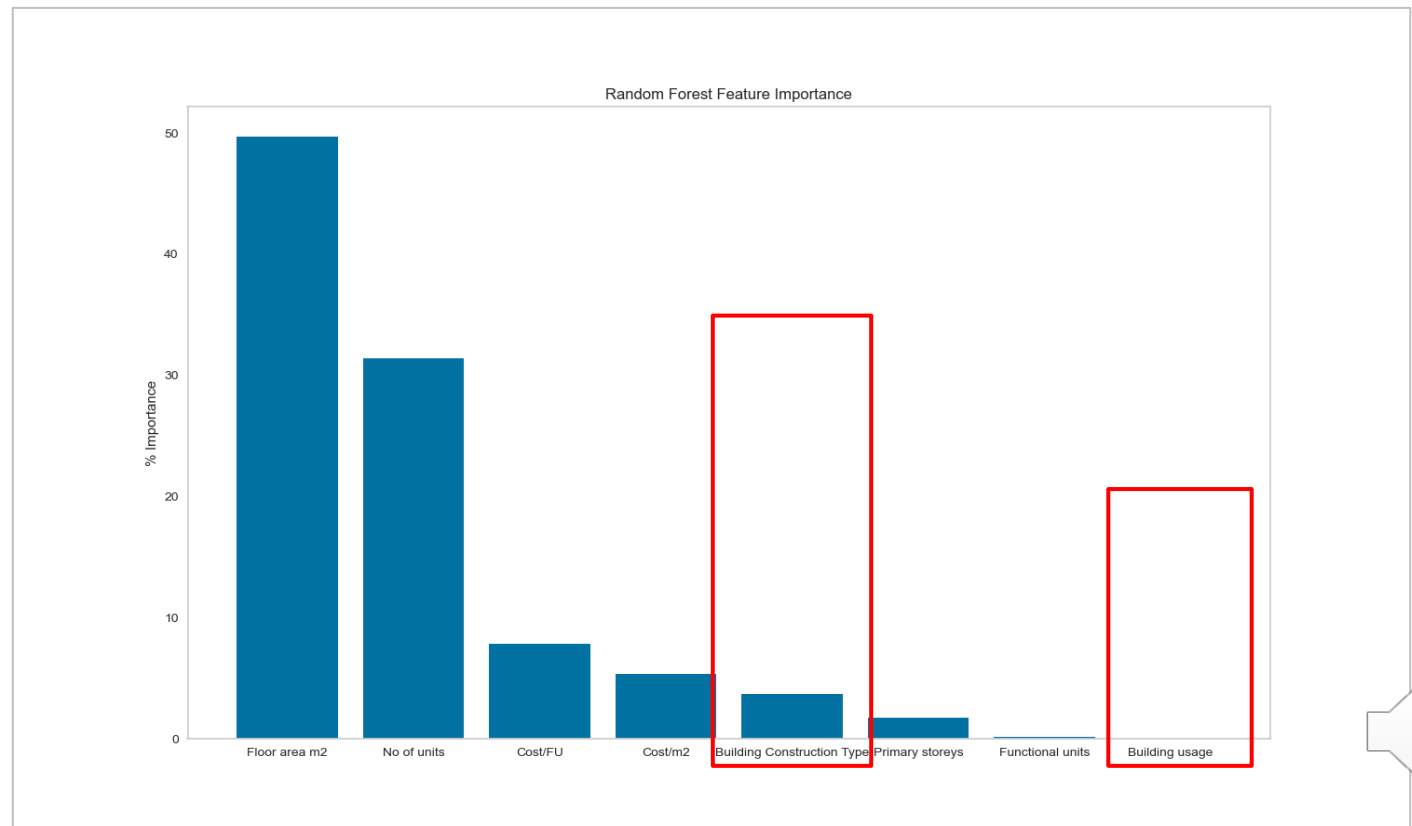
Q1. Would it be possible to include allowances for external factors (location, etc)?

- Location is one factor that could affect cost ,since each location especially when it comes to construction has a different cost (location! location! location!)
- Time series of some economical factors in order to capture the current state of the economy (inflation).
- Number of labors as well as labor wage rates.



Q2 In your opinion,
when comparing
these cost models
which is more
relevant estimation
– building usage or
building
construction type?

- Demonstrates how important are the features used in affecting the output response
- Construction Type has higher impact than the building usage on the response



Q3. How could you make this solution easier for a non-technical person to interact with?

- Visualisation is key! IT provides insights to the user facilitating the decision-making process.
- Interactive dashboards to improve user experience.
- Automating the process in the back-end.

