Springboard Capstone Project 1 - Ideas

Nancy Mao

Ideas

1. Regression
   - use historical house pricing listed on a China real estate brokerage company to predict house value in Beijing based on its location (coordinates)
   - Dataset: https://www.kaggle.com/ruiqurm/lianjia
   - Goal: implement and evaluate different regression models
   - Variables include:

```
Index: 318851 entries, 101084782030 to BJYZ9236355
Data columns (total 25 columns):
url                   318851 non-null object
Lng                   318851 non-null float64
Lat                   318851 non-null float64
Cid                   318851 non-null int64
tradeTime             318851 non-null object
DOM                   160874 non-null float64
followers             318851 non-null int64
totalPrice            318851 non-null float64
price                 318851 non-null int64
square                318851 non-null float64
livingRoom            318851 non-null object
drawingRoom           318851 non-null object
kitchen               318851 non-null int64
bathRoom              318851 non-null object
floor                 318851 non-null object
buildingType          316830 non-null float64
constructionTime      318851 non-null object
renovationCondition   318851 non-null int64
buildingStructure     318851 non-null int64
ladderRatio           318851 non-null float64
elevator              318819 non-null float64
fiveYearsProperty     318819 non-null float64
subway                318819 non-null float64
district              318851 non-null int64
communityAverage      318388 non-null float64
dtypes: float64(11), int64(7), object(7)
```

2. Segmentation:
   - Customer segmentation for e-commerce retailer and customer value analysis
   - Dataset: https://archive.ics.uci.edu/ml/datasets/Online+Retail+II
   - Variables: InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country

3. Forecasting
   - Predict sales for a grocery chain.
   - Goal: predict the sales for individual items at each store location
   - Dataset: https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data
   - Data information:
     - Separate files of information about stores, items, transactions, holiday events are provided.
     - In addition, daily price change information is available as the country where the grocery chain company is located is an oil-dependent country and it's economical health is highly vulnerable to shocks in oil prices.
     - Training and testing datasets are provided separately