

FUTURE SALES PREDICTION

TEAM 22



A photograph of a diverse group of approximately ten people in a modern office or meeting room. Some individuals are seated at a table with laptops, while others stand around, engaged in conversation. The environment is casual and professional, with a mix of men and women wearing various attire like t-shirts, jeans, and button-down shirts. A framed picture hangs on the wall in the background.

NEED OF FUTURE SALES PREDICTION ?

METHODOLOGY



Methodology

Step 1

Importing necessary Libraries
and datasets provided

Step 2

Data Preprocessing

Step 3

Exploratory Data Analysis

Step 4

Model Creation and Testing

Importing libraries and Datasets

```
import numpy as np      # working with array
import pandas as pd     # importing data
from operator import index  #conversion of numeric-obj to interger-obj
from unicodedata import category  #encoding categorial data
import matplotlib.pyplot as plt   #plotting graphs
from scipy import rand    #to find random value to predict
import seaborn as sns    #to create charts and graphs
from sklearn.metrics import mean_squared_error # to calculate mse
```

Provided data description

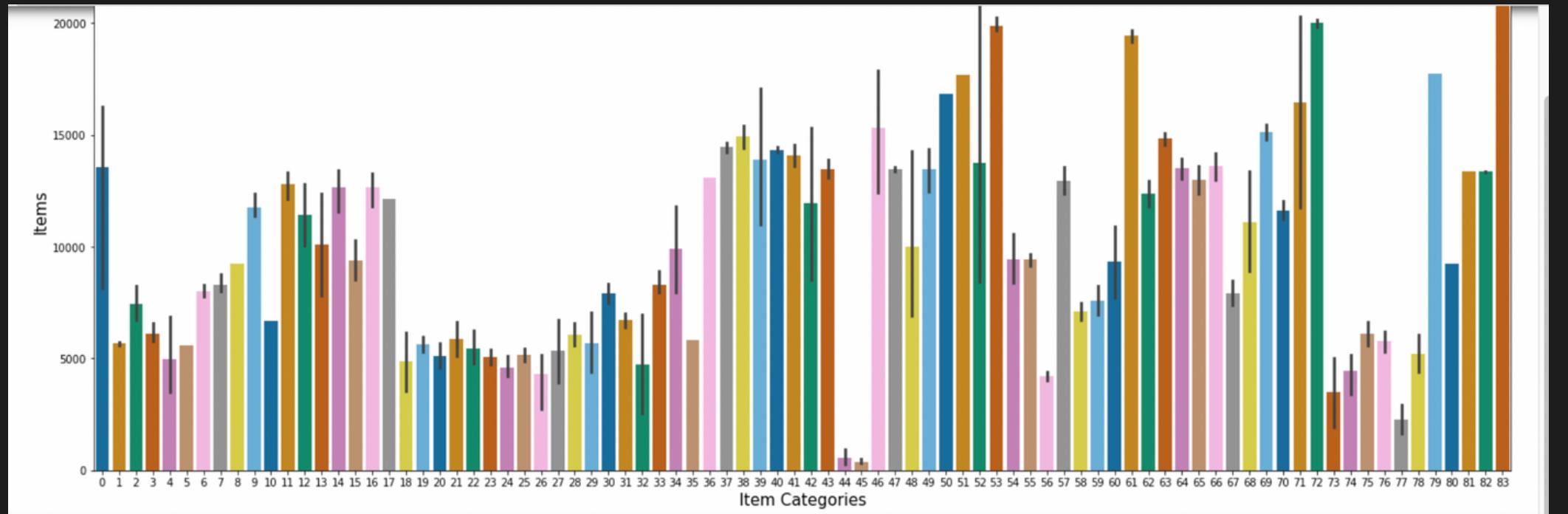
- sales_train.csv - the training set. Daily historical data from January 2013 to October 2015.
- test.csv - the test set. You need to forecast the sales for these shops and products for November 2015.
- sample_submission.csv - a sample submission file in the correct format.
- items.csv - supplemental information about the items/products.
- item_categories.csv - supplemental information about the items categories.
- shops.csv- supplemental information about the shops.

```
train = pd.read_csv('sales_train.csv')
shops = pd.read_csv('shops.csv')
items = pd.read_csv('items.csv')
item_categories = pd.read_csv('item_categories.csv')
```

Data Preprocessing

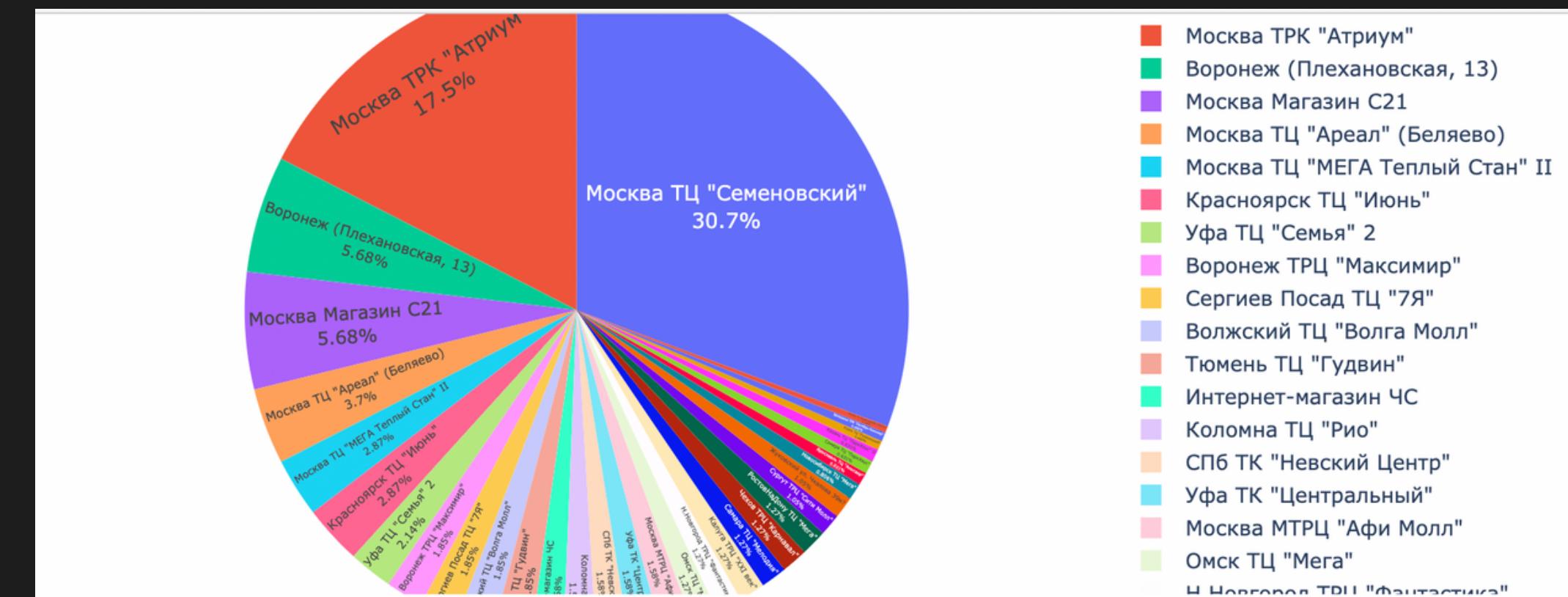
- Merging of two datasets on one primary field.
- Dropping unnecessary columns.

Exploratory Data Analysis



Analysis of number of items sold by each category

Analysis of favourite shops of customers in a particular year



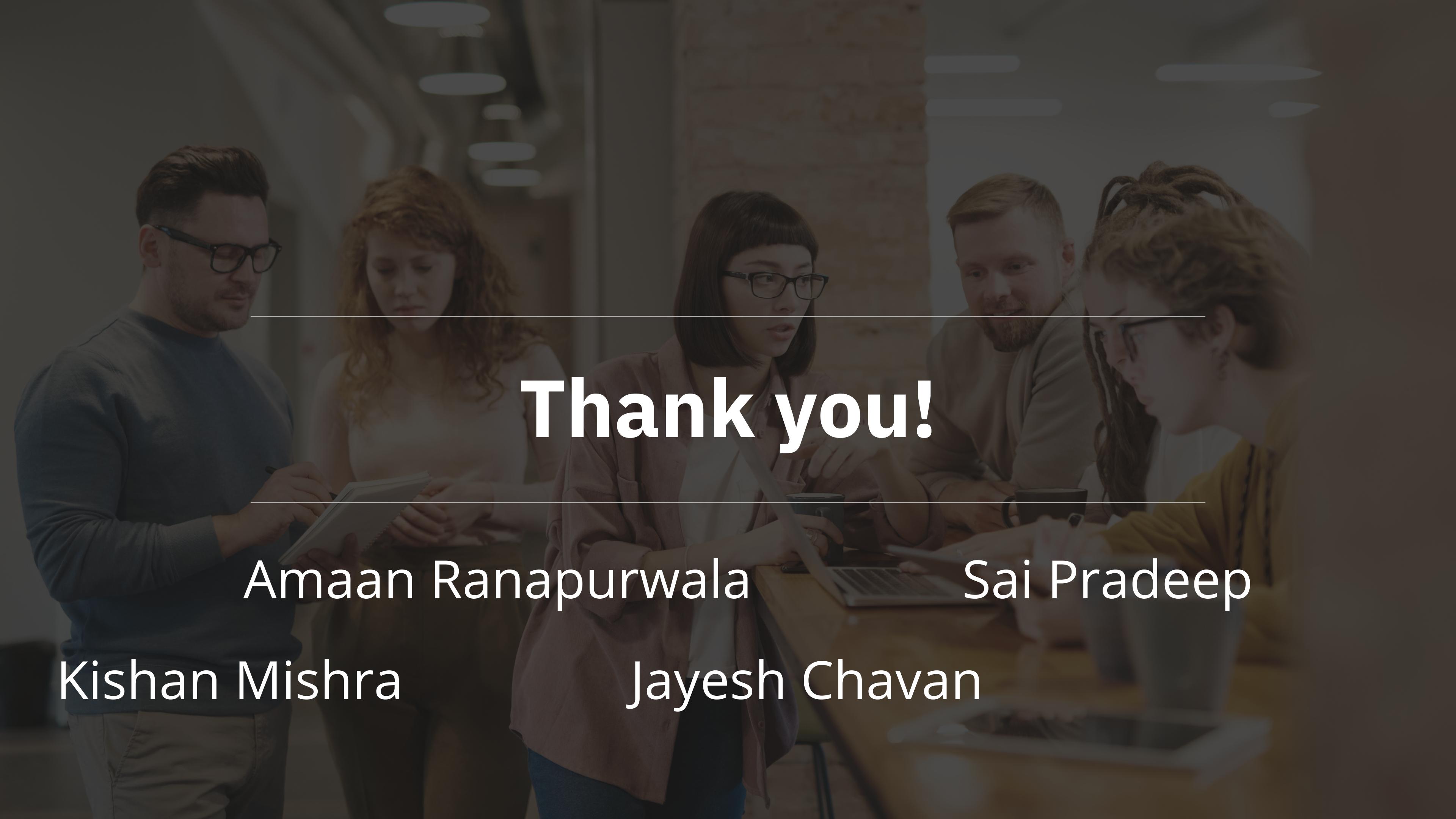
Model Creation and Testing

Random Forest Classifier

```
Training the data....  
Model Accuracy...  
0.42920808764694485  
Model RMSE...  
5.133712843694866  
Predict the test data...  
[2 2 2 ... 1 1 1]
```

XGBoost

```
Training the data....  
Model Accuracy...  
0.6780222442647269  
Model RMSE...  
3.8557210068557577  
Predicting the test data.  
[2 2 2 ... 1 1 1]
```

A group of five people are gathered around a table in a modern office environment. A man on the far left wears glasses and a dark turtleneck, writing in a notebook. Next to him is a woman with long curly hair in a light-colored top. In the center, a woman with short dark hair and glasses looks intently at a laptop screen. To her right, a man with a beard and a light-colored shirt also looks at the screen. On the far right, a woman with braided hair and glasses is partially visible, looking down at something on the table.

Thank you!

Amaan Ranapurwala

Kishan Mishra

Sai Pradeep

Jayesh Chavan