Store Sales Prediction

Snehaa Durairaj

18th March 2024


BUT WHY????

Importance of Sales Forecasting

1. Enhancing Planning

Sales forecasting provides critical insights for inventory management, resource allocation, and strategic decision-making, enabling businesses to plan effectively for future demand.

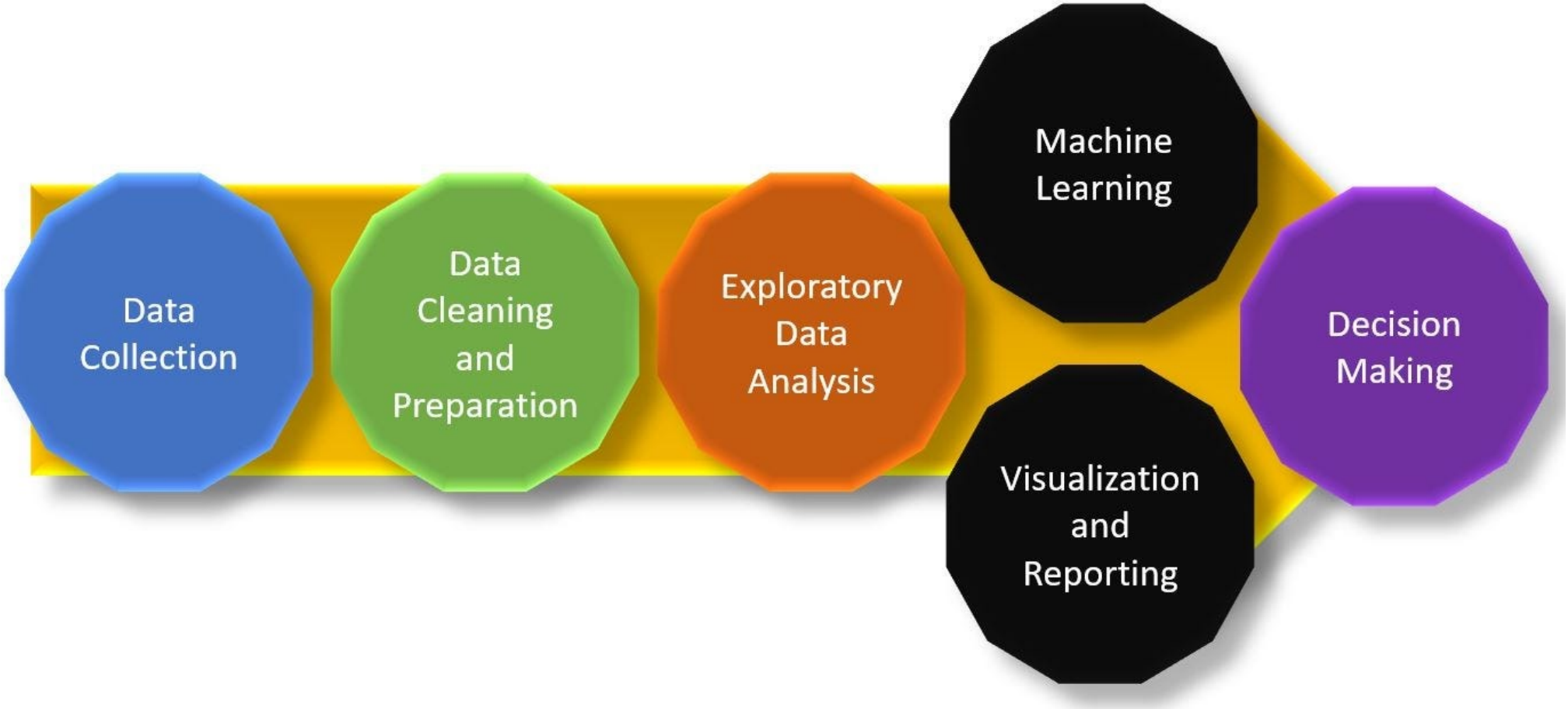
2. Market Analysis

By analyzing historical sales data, businesses can gain a deeper understanding of market trends, customer behavior, and the impact of external factors on sales performance.

3. Financial Management

Accurate sales predictions are essential for budgeting, financial planning, and setting realistic revenue targets, contributing to the overall financial health of the organization.

Overview


Data Collection
Data Cleaning and Preparation
Exploratory Data Analysis
Machine Learning
Visualization and Reporting
Decision Making

Overview of the Dataset

Tables :

1. Stores
2. Sales
3. Transactions

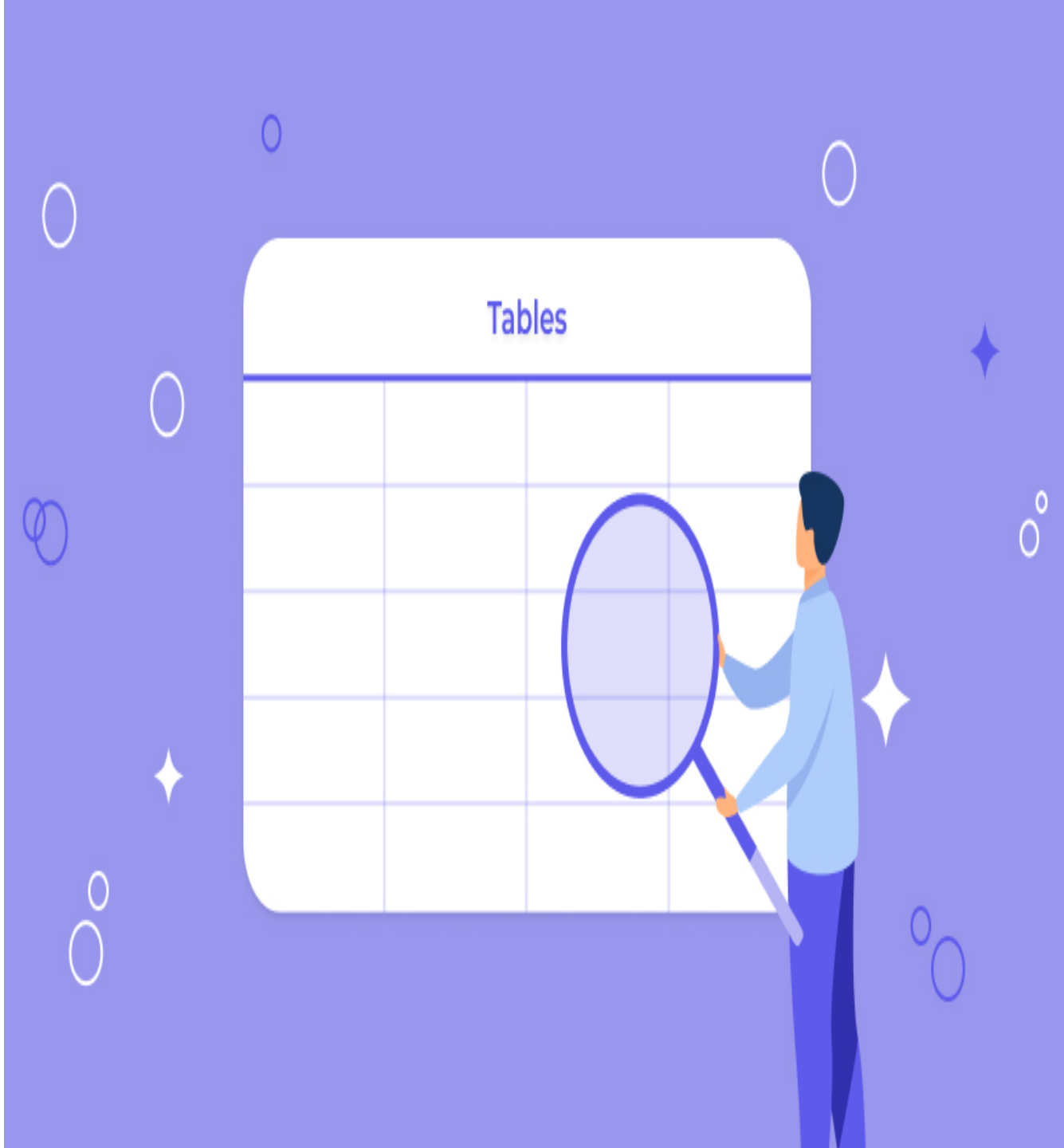
Data Span from 2013 to 2017

Target Variable: Sales

Stores
Store_nbr
City
State
Type
Cluster

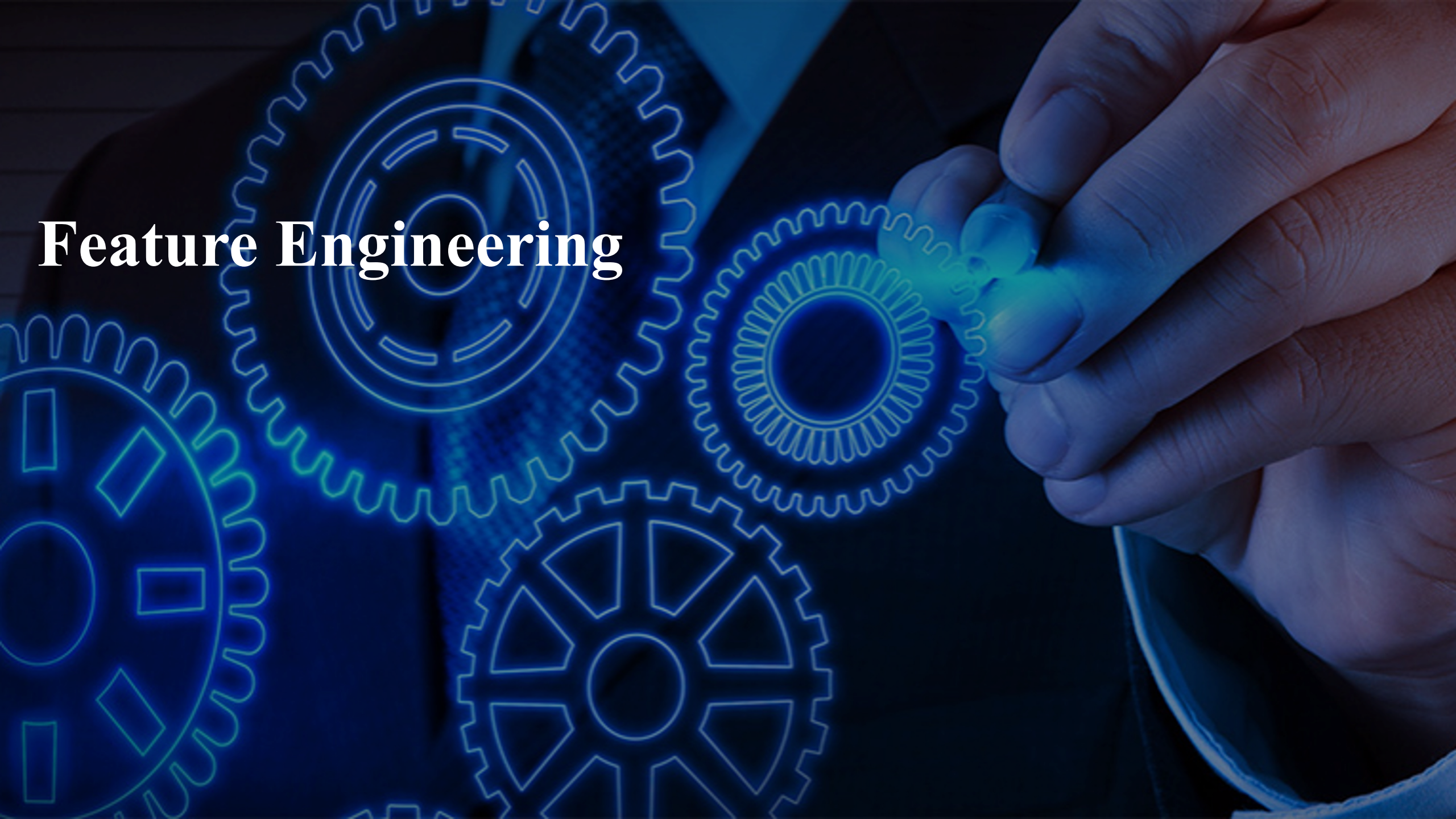
Sales
Date
Store_nbr
Family
Sales
OnPromotion

Transactions
Date
Store_nbr
Transactions

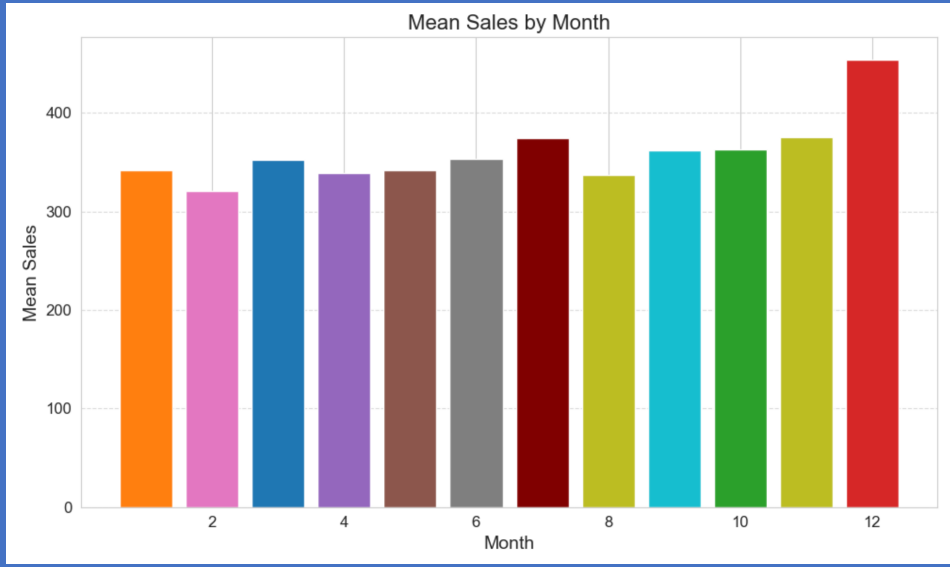


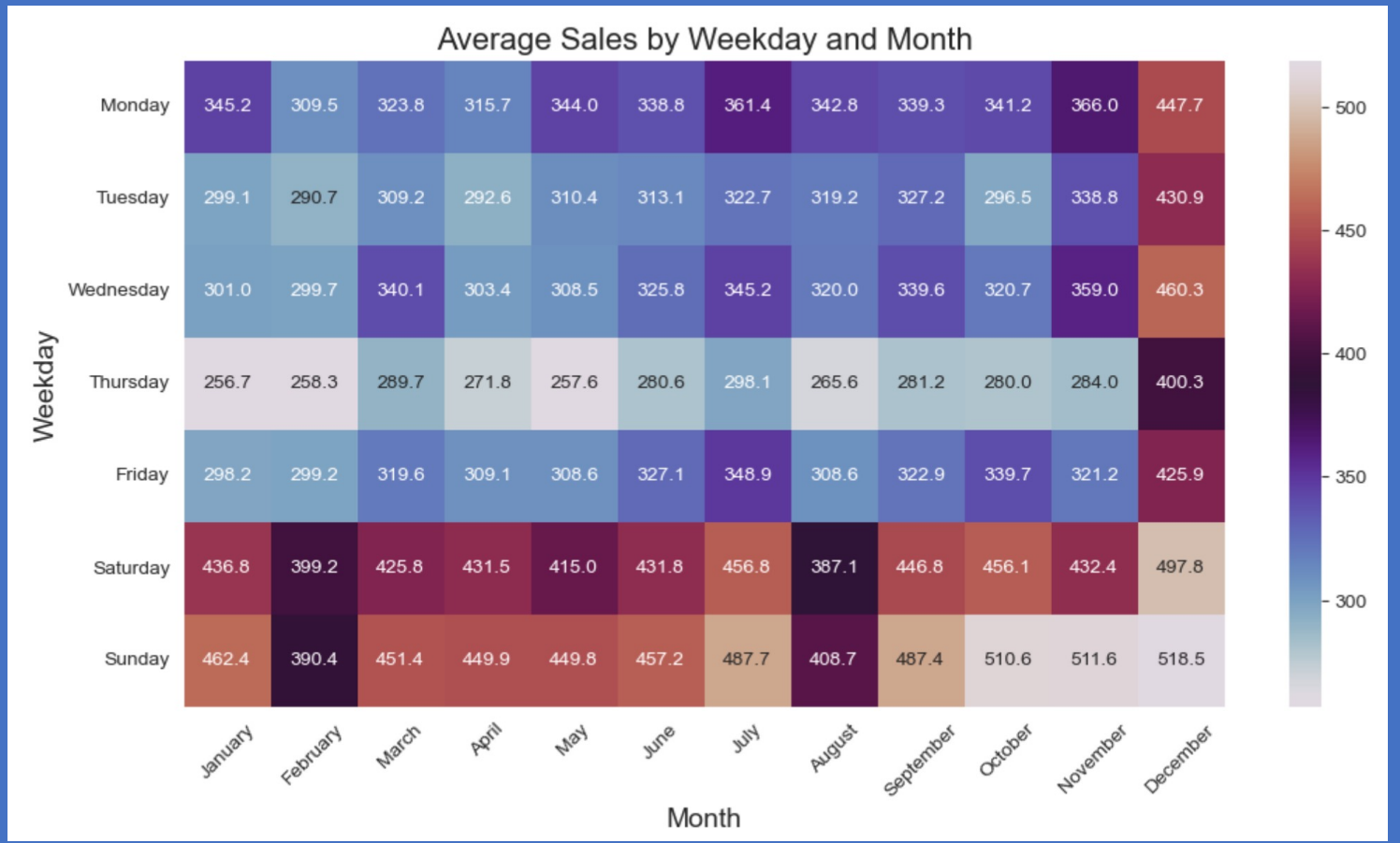

And
Preparation

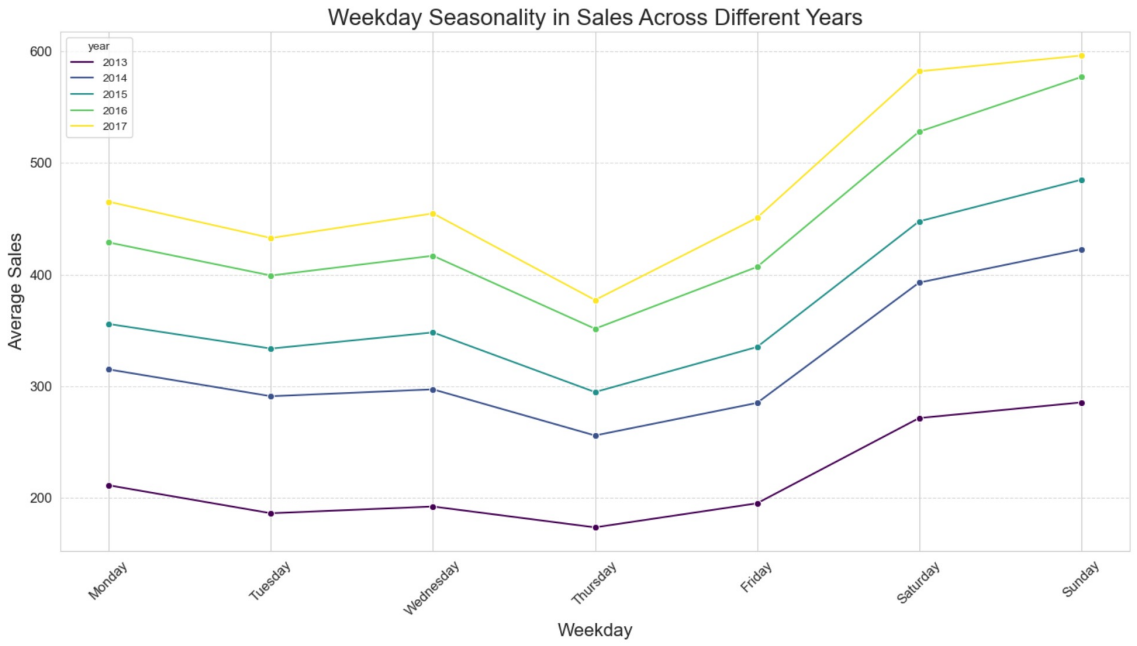
Feature Engineering

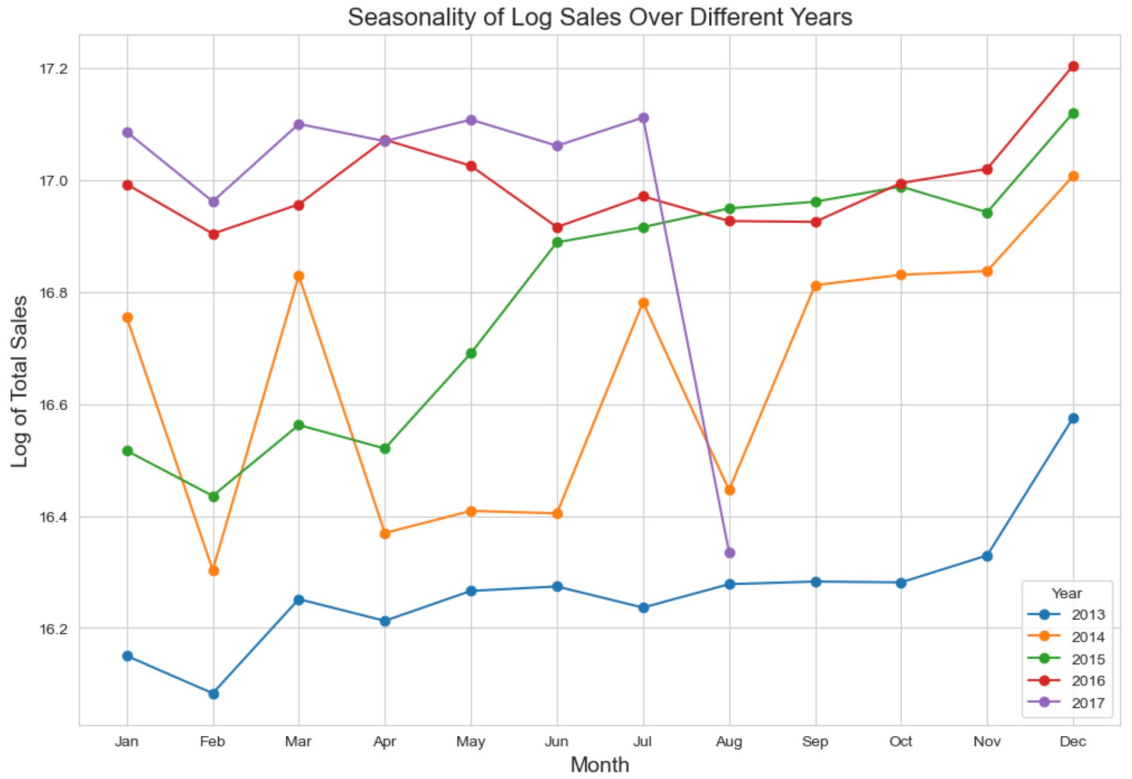


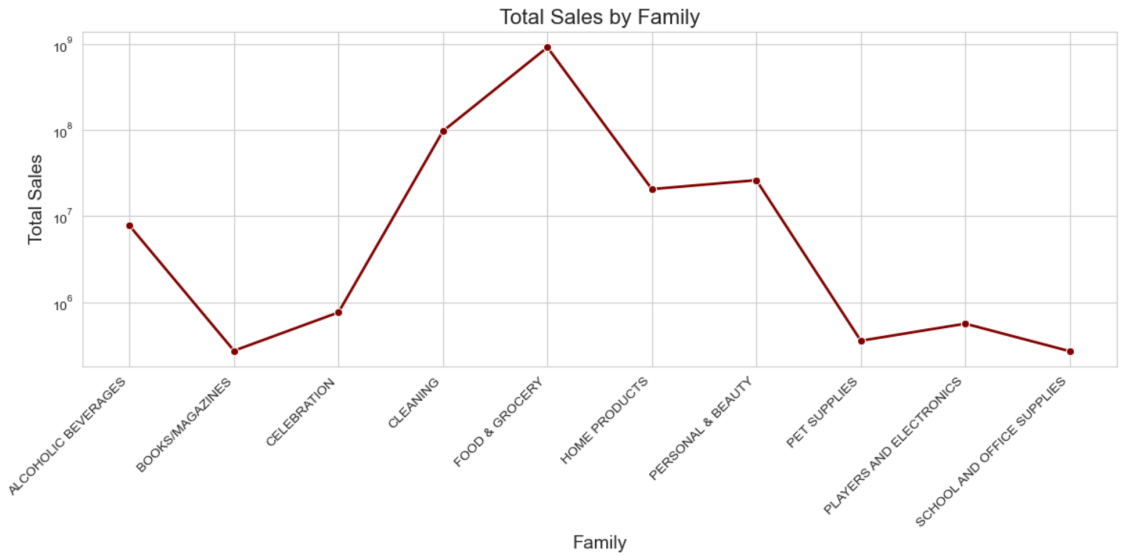
DETECTING OUTLINERS!!!

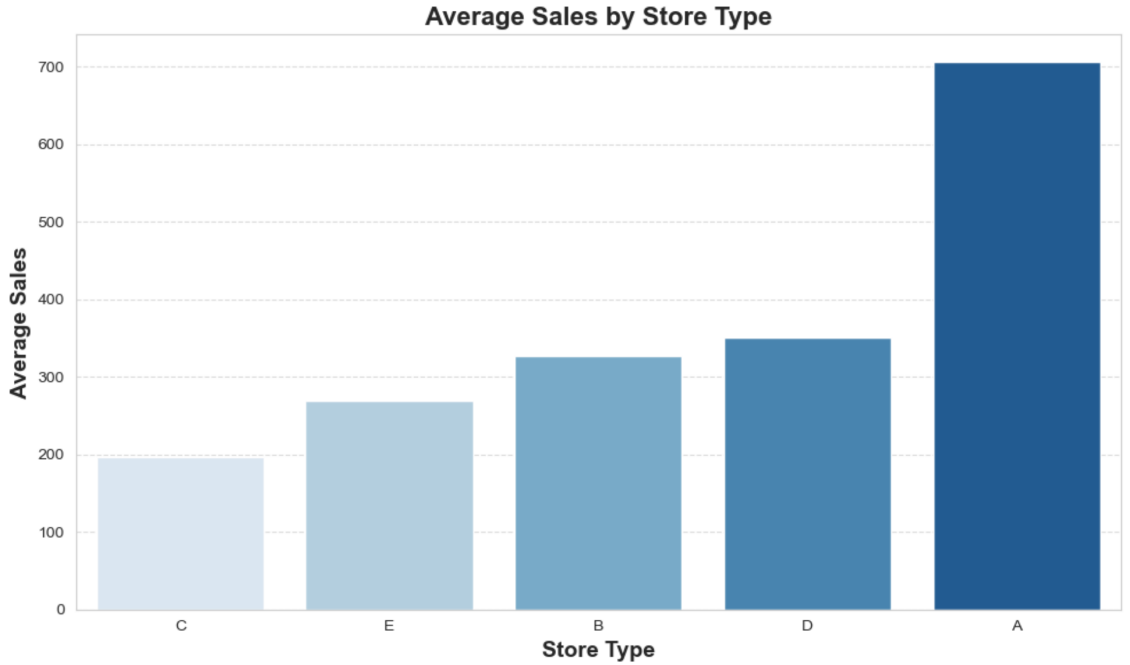

Mean Sales by Month
Mean Sales
0
100
200
300
400
2
4
6
8
10
12
Month


Average Sales by Weekday and Month
Weekday
Monday 345.2 309.5 323.8 315.7 344.0 338.8 361.4 342.8 339.3 341.2 366.0 447.7
Tuesday 299.1 290.7 309.2 292.6 310.4 313.1 322.7 319.2 327.2 296.5 338.8 430.9
Wednesday 301.0 299.7 340.1 303.4 308.5 325.8 345.2 320.0 339.6 320.7 359.0 460.3
Thursday 256.7 258.3 289.7 271.8 257.6 280.6 298.1 265.6 281.2 280.0 284.0 400.3
Friday 298.2 299.2 319.6 309.1 308.6 327.1 348.9 308.6 322.9 339.7 321.2 425.9
Saturday 436.8 399.2 425.8 431.5 415.0 431.8 456.8 387.1 446.8 456.1 432.4 497.8
Sunday 462.4 390.4 451.4 449.9 449.8 457.2 487.7 408.7 487.4 510.6 511.6 518.5
January
February
March
April
May
June
July
August
September
October
November
December
Month
500
450
400
350
300


Weekday Seasonality in Sales Across Different Years
year
2013
2014
2015
2016
2017
600
500
400
300
200
Average Sales
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Weekday


Seasonality of Log Sales Over Different Years
17.2
17.0
16.8
16.6
16.4
16.2
Log of Total Sales
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Year
2013
2014
2015
2016
2017


Total Sales by Family
Total Sales
10^9
10^8
10^7
10^6
ALCOHOLIC BEVERAGES
BOOKS/MAGAZINES
CELEBRATION
CLEANING
FOOD & GROCERY
HOME PRODUCTS
PERSONAL & BEAUTY
PET SUPPLIES
PLAYERS AND ELECTRONICS
SCHOOL AND OFFICE SUPPLIES
Family


Average Sales by Store Type
Average Sales
700
600
500
400
300
200
100
0
C
E
B
D
A
Store Type


Model Building

Decision Tree - From Root to Leaf: Clear Decision Making with Trees

Without Feature Selection:
11 Features

RMSE: 0.80

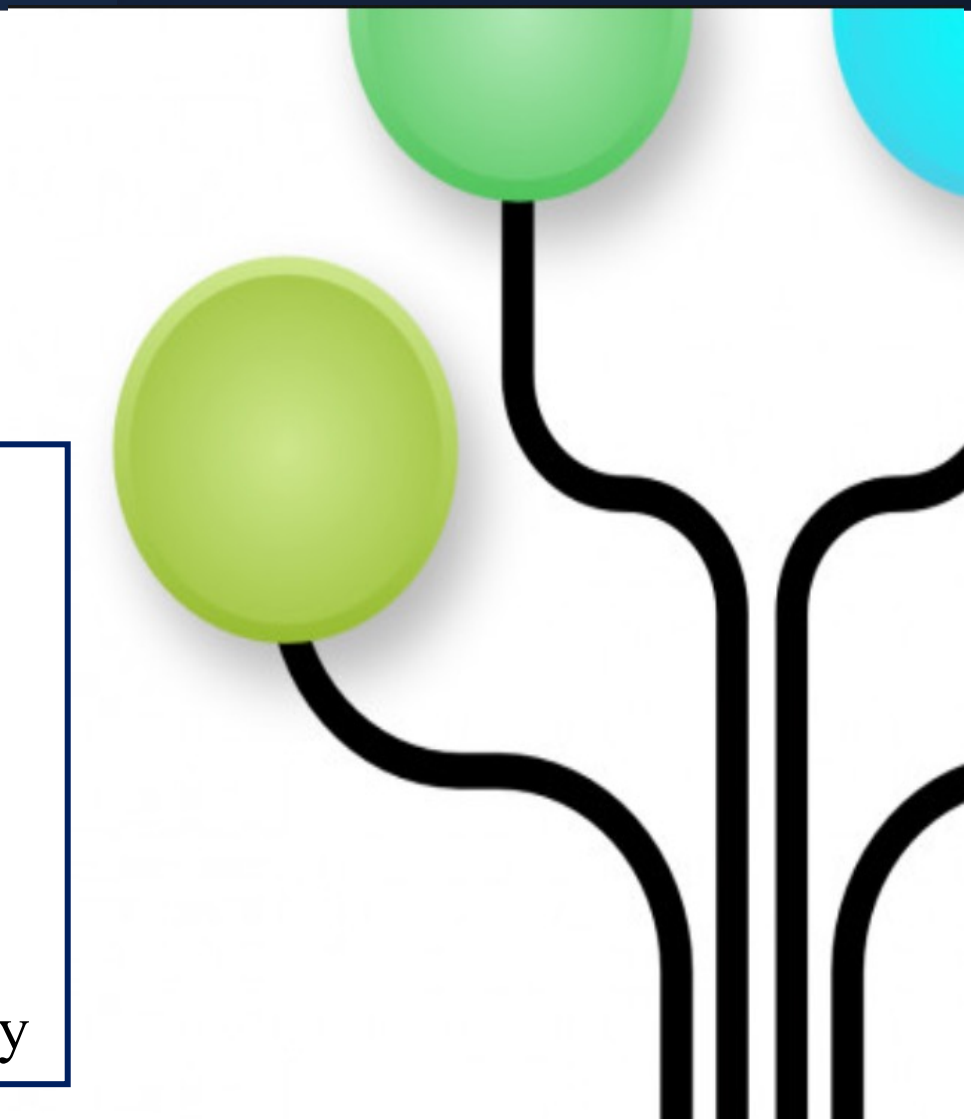
Var: 0.36

With Feature Selection: 3 Features

RMSE: 0.86

Var: 0.17

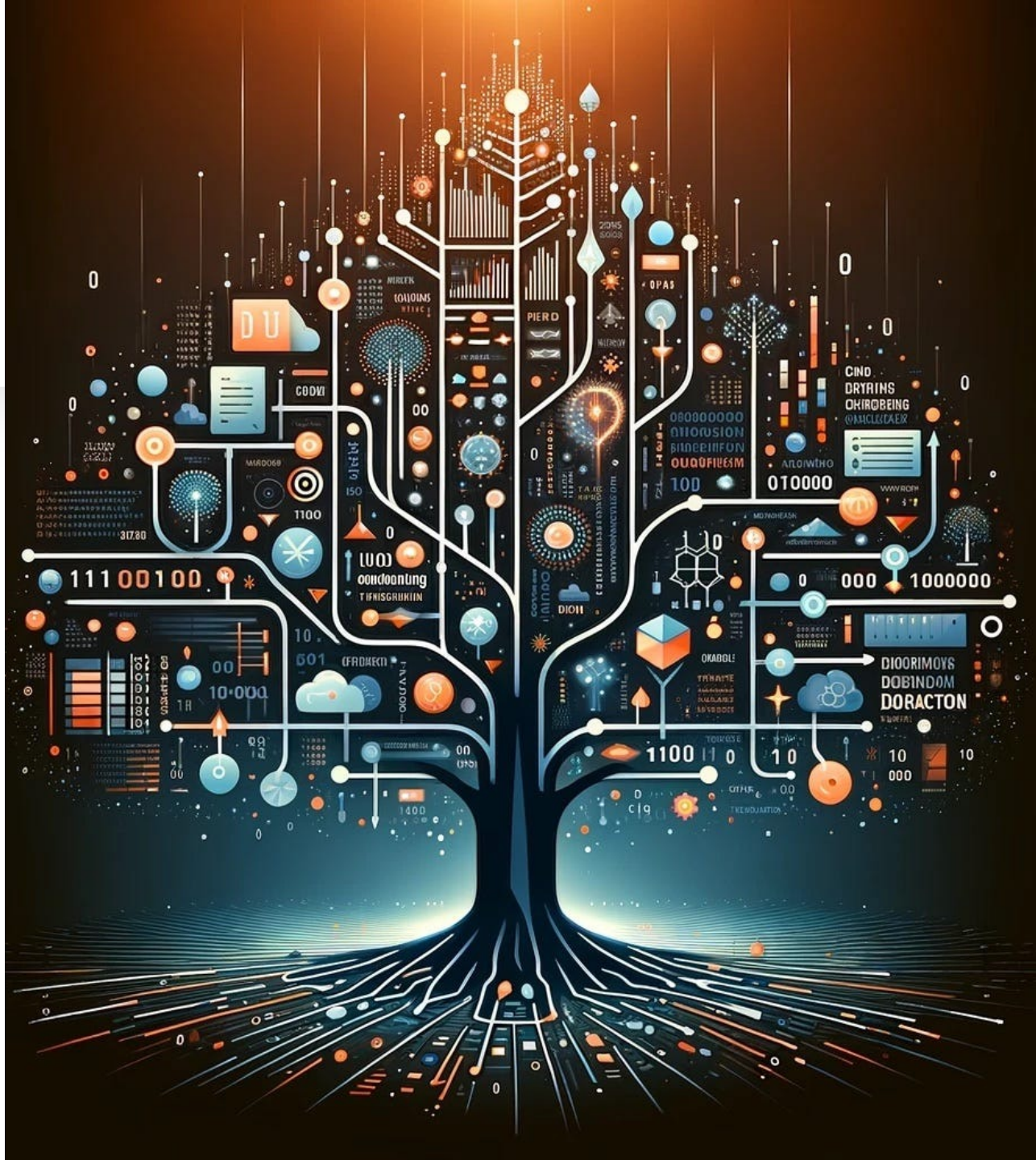
Features Selected:
OnPromotion, Family, Day



Random Forest- A Forest of Decision Trees for Enhanced Predictions

- Without Feature Selection – 11 Features
- RMSE: 0.70
- Var: 0.46

- With Feature Selection: - 3 Features
- RMSE: 0.69
- Var: 0.47
- Features Selected: OnPromotion, Family, Day



Boosting Methods

Gradient Boosting Method

Without Feature Selection – 11 Features

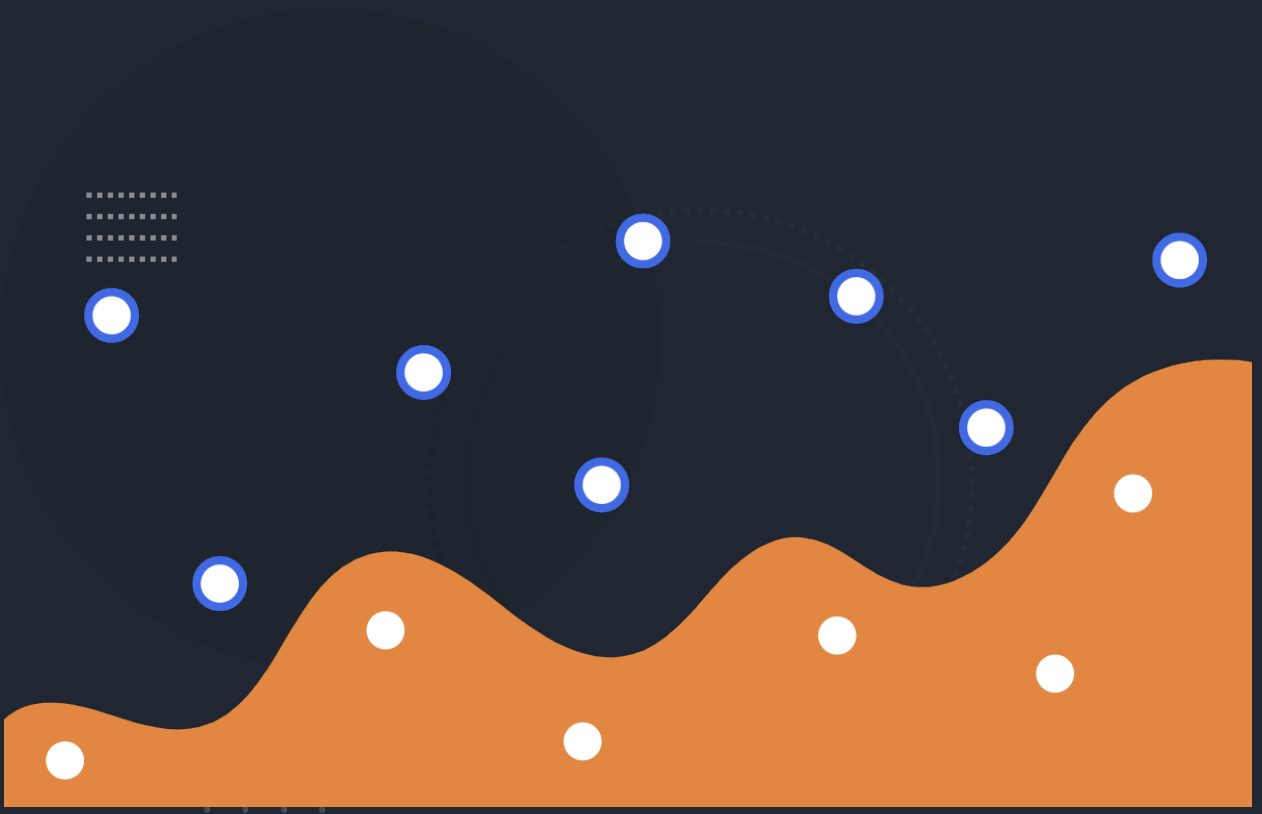
RMSE: 0.74
Var: 0.41

ADA Boosting Method

Without Feature Selection – 11 Features

RMSE: 0.86
Var: 0.18

SVM : Finding the Optimal Separation



Model	RMSE	Explained Var
SVM with ‘rbf’	0.88	0.15
SVM with ‘linear’	0.90	0.10
SVM with ‘sigmoid’	0.91	0.09

Model	RMSE	Explained Variance
Decision Tree	0.80	0.36
Decision Tree with Wrapper Select	0.86	0.17
Random Forest	0.70	0.46
Random Forest with Wrapper Select	0.69	0.47
Gradient Boosting	0.74	0.41
ADA Boosting	0.86	0.18
SVM with ‘rbf’ kernel	0.88	0.15
SVM with ‘linear’ kernel	0.90	0.10
SVM with ‘sigmoid’ kernel	0.91	0.09


THANK YOU!