Iowa Liquor Project

Group 4:

Zhang Huiyan 21436576
Zheng Weina 21426716
Liang Jinbing 21419728
Yang Ziqing 21462283

CONTENTS

01 | **Business Understanding**
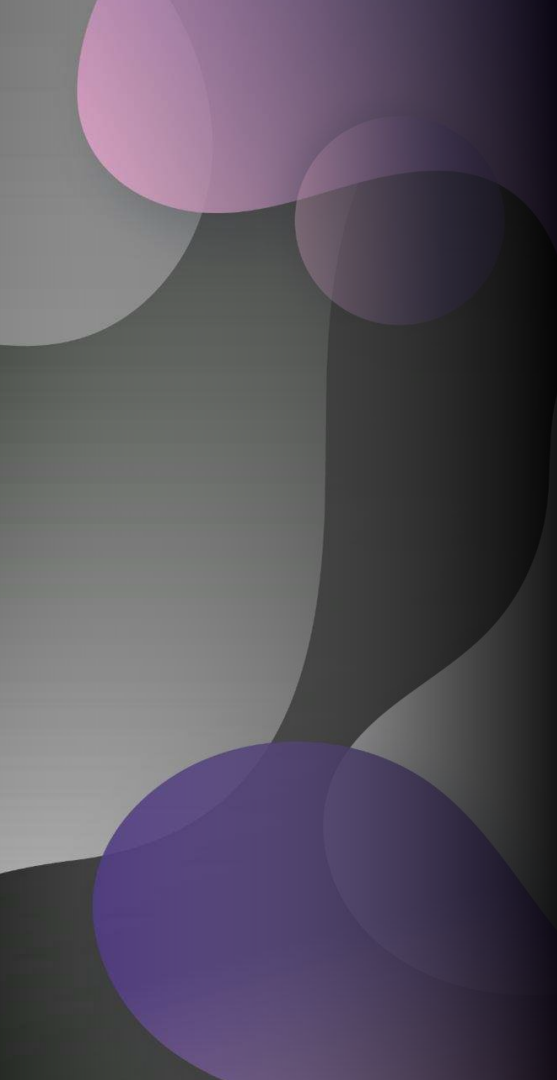
02 | **Data Preprocessing**

03 | **Modeling**

04 | **Conclusions**

01 | **Business Understanding**

About Cheers Alcohol

Product Lines

- Wines 🍷

- Beers 🍻

- Ciders

- Spirits & Liqueur

Ref:https://www.cheerstheliquorshop.co.uk/

Iowa Liquor Market Trend
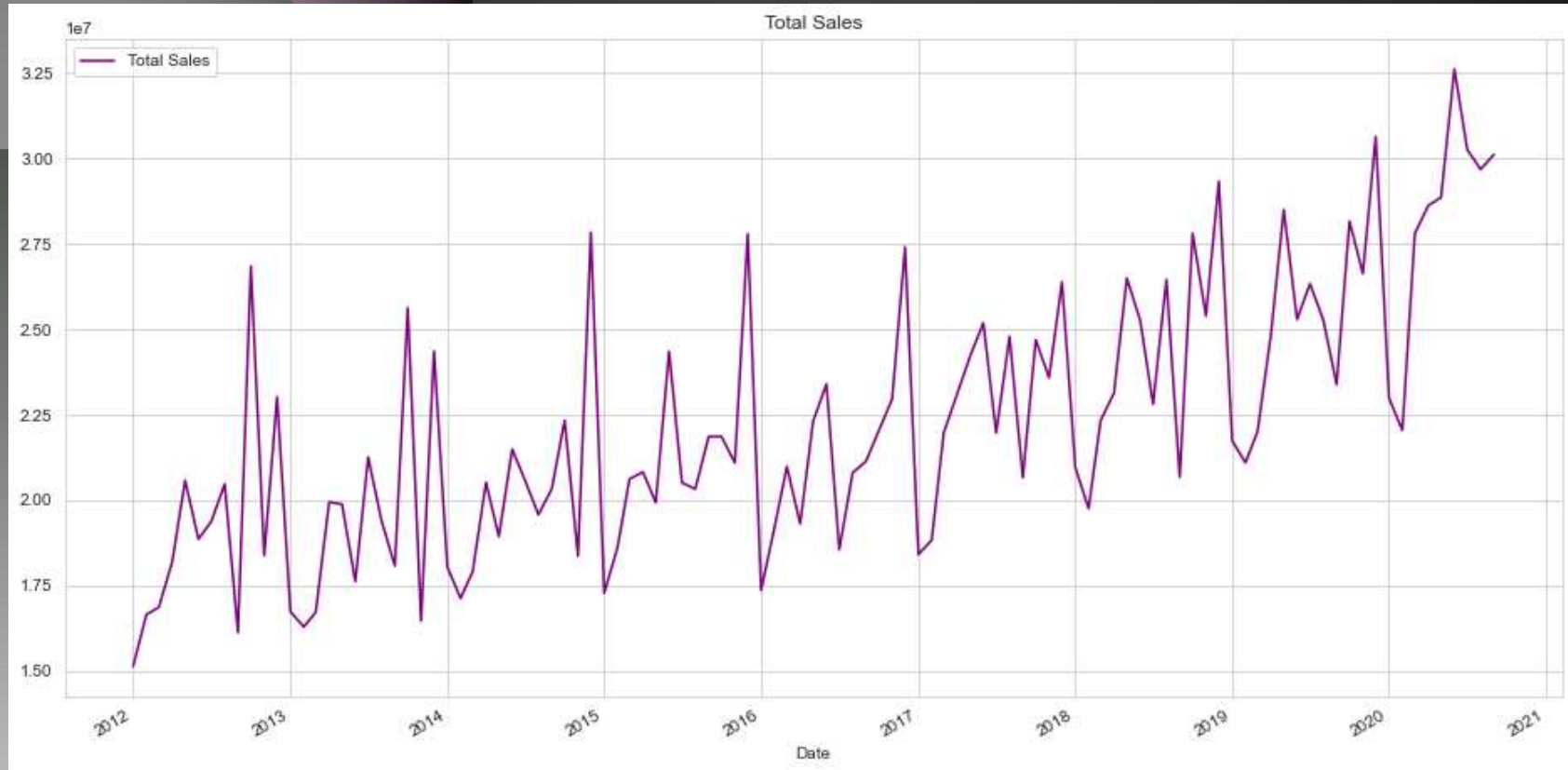
STRATEGIC OBJECTIVES

If Cheers Alcohol wants to launch 3 stores, which county/counties should we target?

Predict the total sales for the counties we selected to launch stores in each month of 2021.

What items would we sell in each store? Select **top 10 items** in each store to promote.

02

Data

Processing

Main processing

Features selection

Drop features

Target feature

Cleaning

Missing data

Duplicated data

Outliers

Transformation

Reduce data size

Subsets for modeling

Feature Selection

Drop Features

- Invoice/Item Number
- Store Name
- Address, Zip Code, Store Location
- Item Description
- Vendor Name
- Volume Sold(Gallons)

Target Feature

- Sales (Dollars)

Transformation

Date

- Object → datetime type
- Change the datetime format "%Y-%m"
- Set date as index

Reduce data memory size

- Int64 → int32/int8
- Float64 → float32
- Remove capitalization

3.5+G → 1.5+G

Cleaning Data

Missing Value

- Address
- City
- Zip Code
- Store Location
- County Number/ County
- Category/ Category Name
- Vendor Number/ Name
- State Bottle Cost
- State Bottle Retail
- Sale(Dollars)

Duplicated Data

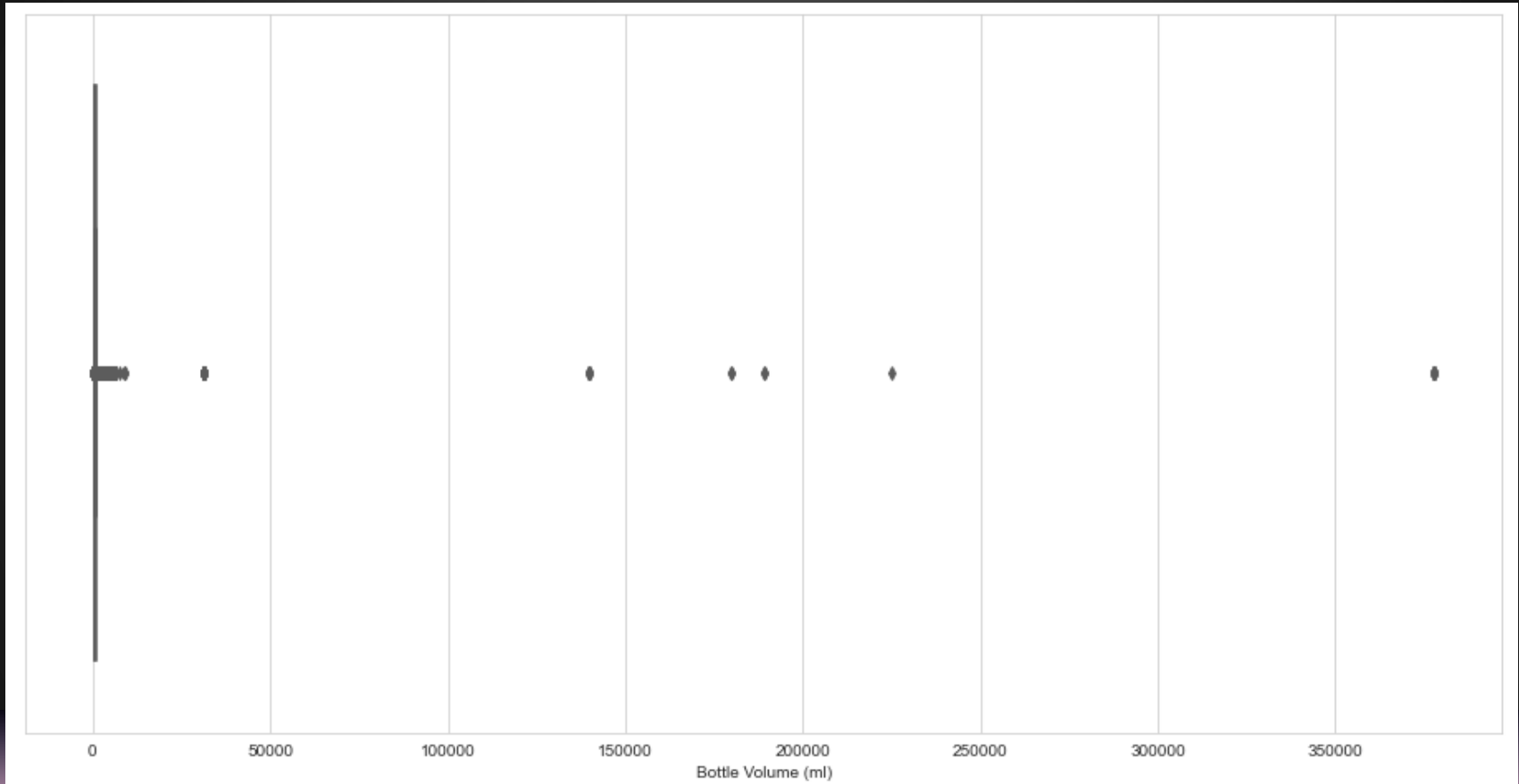- No duplicate data

Cleaning Data

Outliers

- Bottle Volume(ml)
- State Bottle Cost
- State Bottle Retail
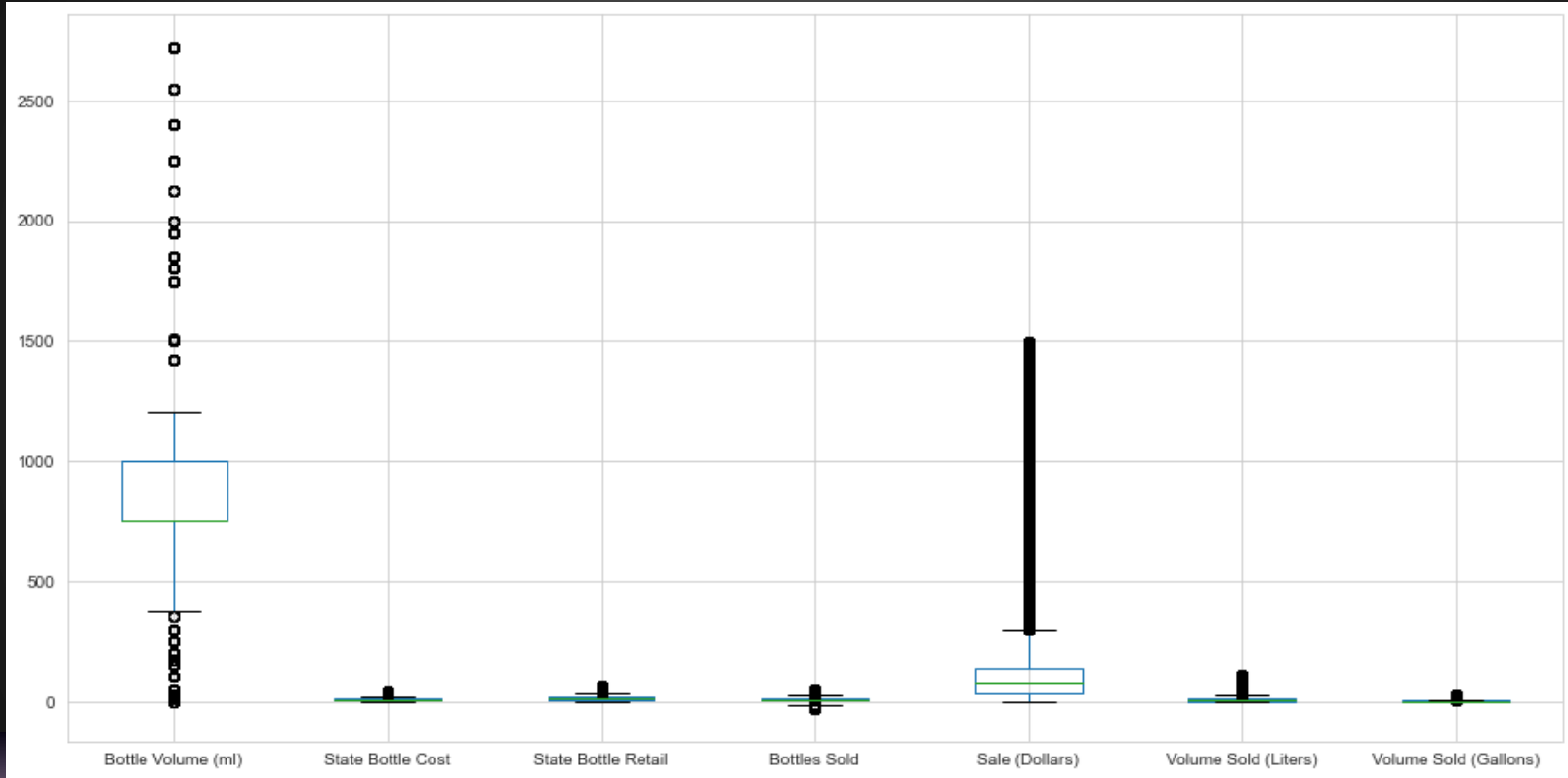- Bottles Sold
- Sales (Dollars)
- Volume Sold (Liters)

Treatment

- Replace outlier with 3 std

Boxplot of bottle volume(ml) with outliers

Boxplot of numeric column after replacing outliers

How to choose the new store locations?

Which counties sold the most alcohol?

Total Sales of county

Are these counties already overstocked?

Number of stores in the county

Which counties sold most liquor?

01 Sum up the total sales of each store

02 Take average of store level sales by each county

03 Rank the top 10 counties

Top 10 counties by average total sales

1	Dickinson	6	johnson
2	Polk	7	linn
3	cerro gordo	8	story
4	scott	9	jefferson
5	woodbury	10	pottawattamie

Top 10 Sales Counties

Top 10 Counties by Average Total Sales

Word cloud for Top 10 Counties by Sales

County Name Wordcloud
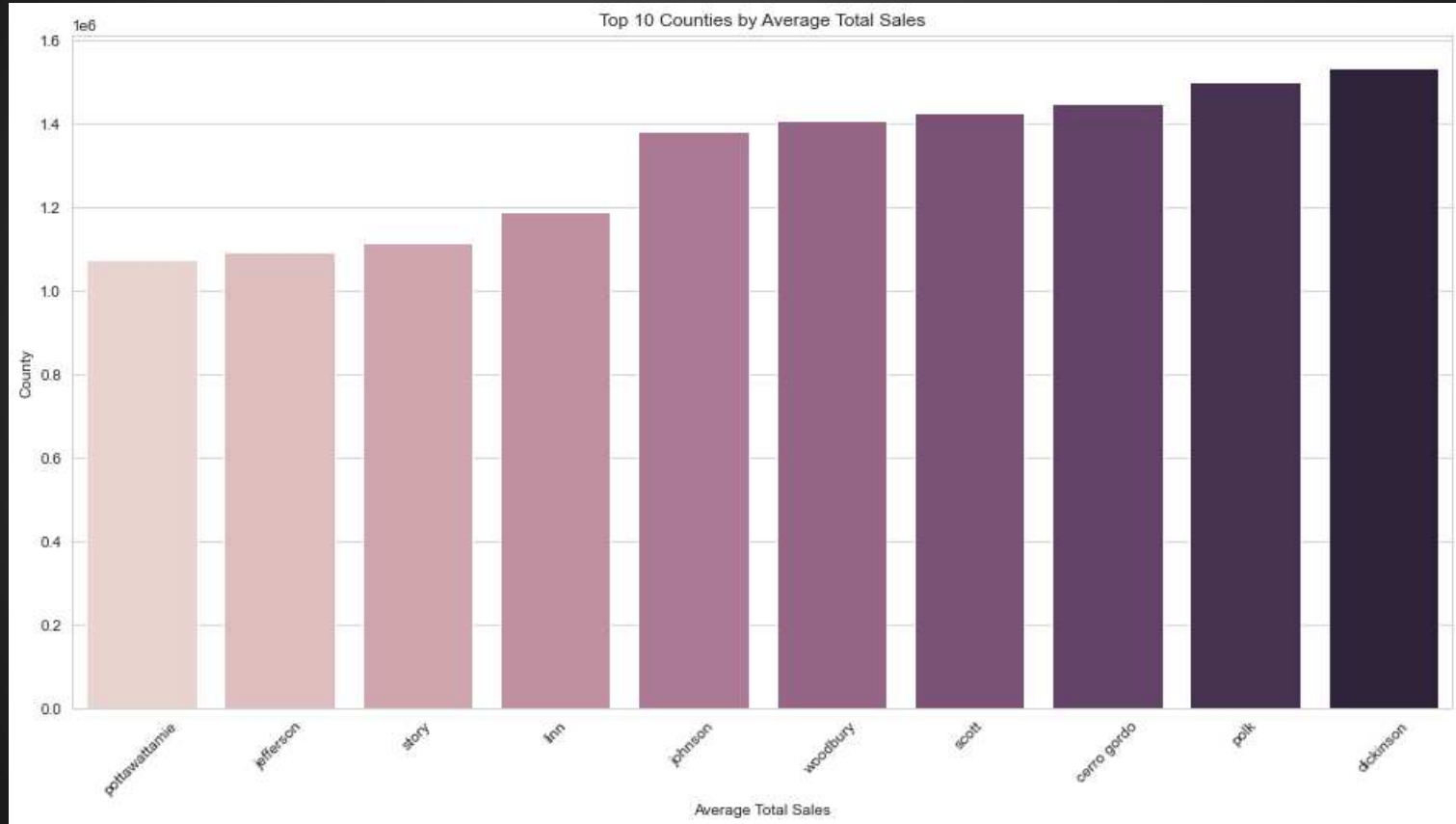
Do these counties already have too much store?

01 Select the distinct Store Number in each county
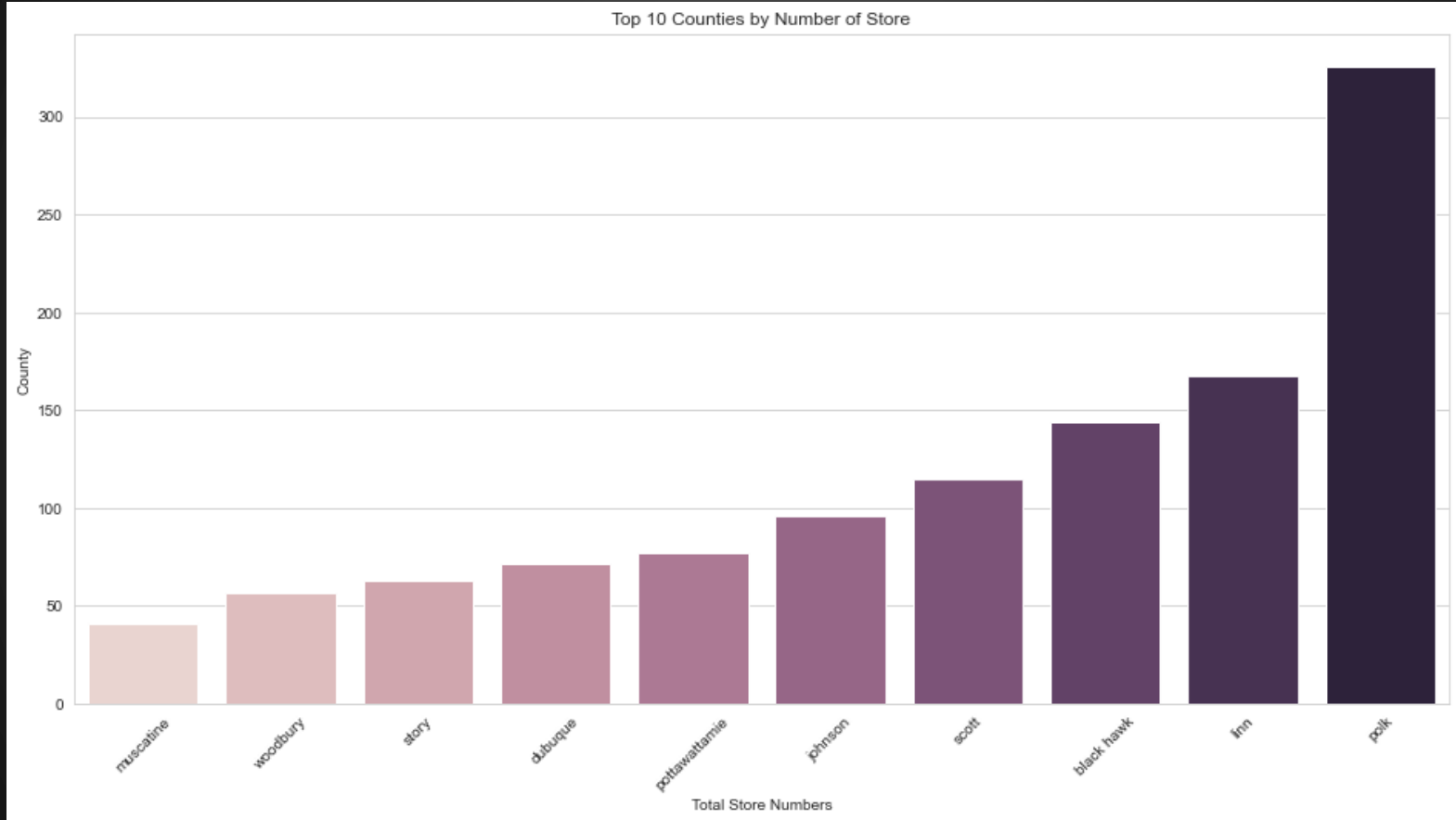
02 Count the Stores in each county

03 Rank the top 10 counties with most stores

Top 10 counties by number of stores

1	Polk	6	pottawattamie
2	Linn	7	dubuque
3	Black hawk	8	Story
4	Scott	9	woodbury
5	Johnson	10	muscatine

Top 10 Counties by Store Numbers

Top 10 Counties by Number of Store

Word cloud for Top 10 Counties by Sales

County Name Wordcloud

black muscatine linn scottpolk story hawk dubuque johnson woodbury pottawattamie
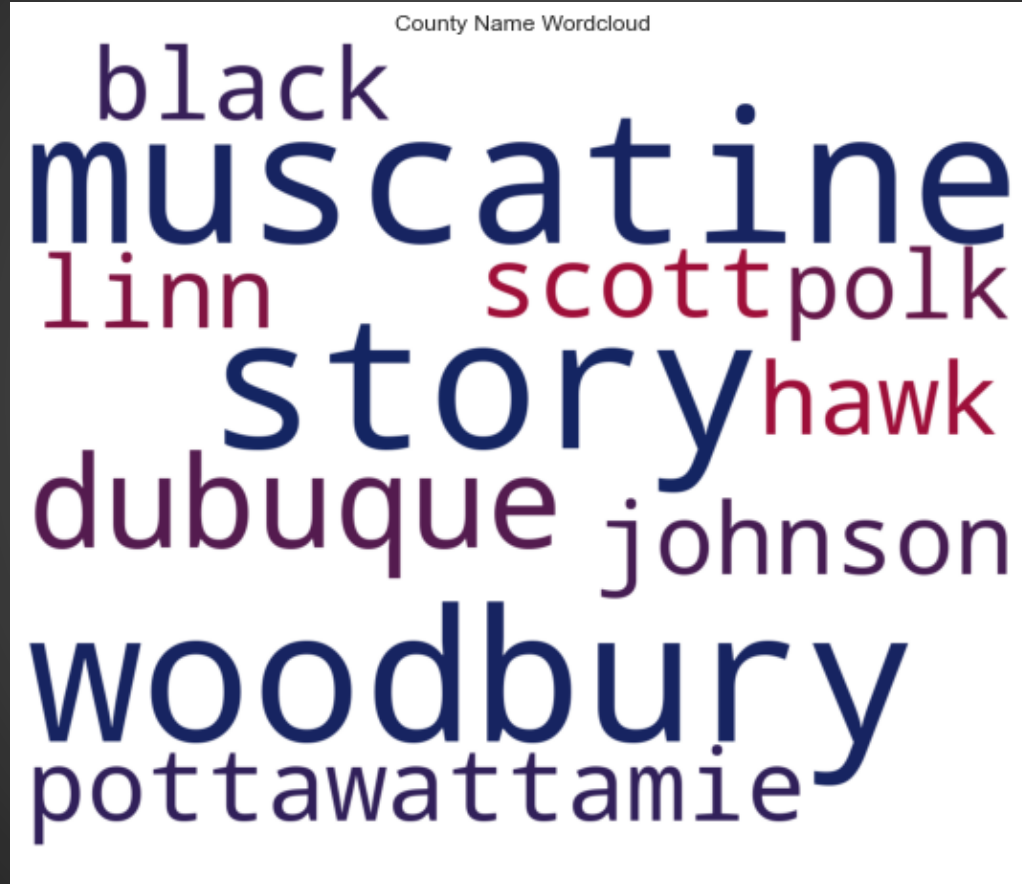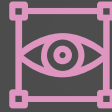
We choose these counties…

In Top 10 counties by average total sales
NOT in Top 10 counties by number of stores

Dickinson   **Jefferson**   **Cerro Gordo**

10 items we choose to sell in Dickinson

01 Select the county we will to locate

02 Sum up the sales by item number

03 Rank the top 10 items

Top 10 items by total sales in Dickinson

1	Captain Morgan Spiced Rum	6	Titos Handmade Vodka
2	Black Velvet	7	Barton Vodka
3	Crown Royal Canadian Whisky	8	Grey Goose Vodka
4	Jack Daniels Old #7 Black Lbl	9	Malibu Coconut Rum
5	Fireball Cinnamon Whiskey	10	Titos Handmade Vodka

10 items we choose to sell in Jefferson/Cerro Gordo

Top 10 items by total sales in Jefferson

1. Black Velvet
2. Hawkeye Vodka
3. Jack Daniels Old #7 Black Lbl
4. Templeton Rye
5. Titos Vodka
6. Captain Morgan Original Spiced
7. Crown Royal
8. Captain Morgan Spiced Rum
9. Fireball Cinnamon Whiskey
10. Jameson

Top 10 items by total sales in Cerro Gordo

1. Black Velvet
2. Captain Morgan Spiced Rum
3. Hawkeye Vodka
4. Fireball Cinnamon Whiskey Mini Dispenser
5. Crown Royal Canadian Whisky
6. Jack Daniels Old #7 Black Lbl
7. Captain Morgan Spiced Barrel
8. Absolut Swedish Vodka 80 Prf
9. Titos Handmade Vodka
10. Black Velvet

03 | **Modelling**

Training data: 0.75
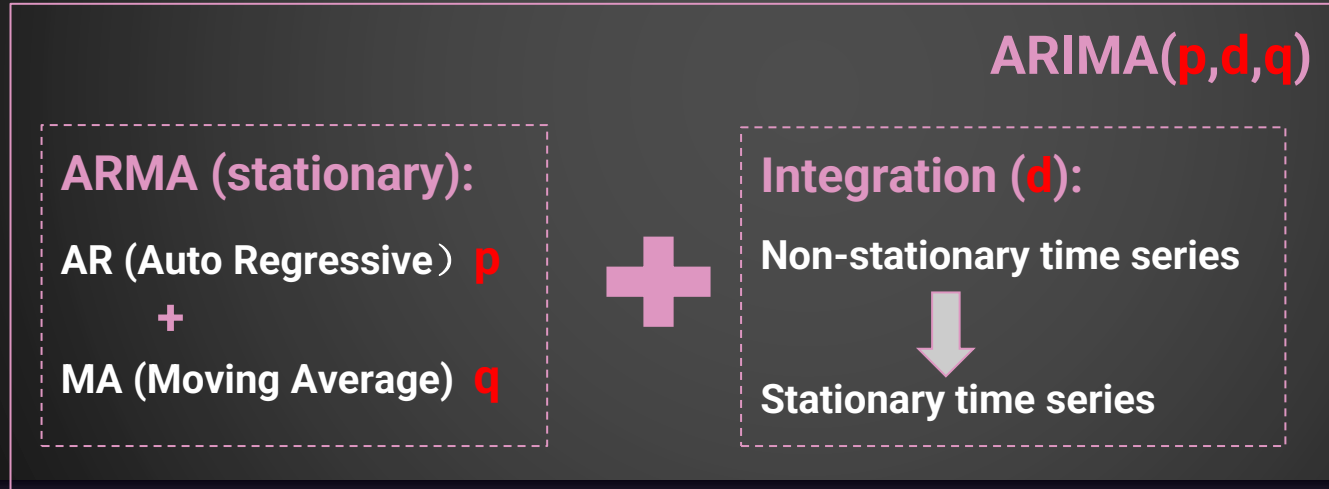**Test data: 0.25**

ARIMA

SARIMAX

Facebook Prophet

ARIMA (Autoregressive integrated moving average)

- A common time series forecasting model
- Predict the future values based on the past values
- An extension of simpler ARMA model (Autoregressive Moving Average)

ARIMA(p,d,q)

ARMA (stationary):

AR (Auto Regressive）p

        +

MA (Moving Average) q

➕

Integration (d):

Non-stationary time series

⬇

Stationary time series

SARIMAX (Seasonal ARIMA with eXogenous factors)

- An updated version of the ARIMA model
- Include seasonal effects and eXogenous factors

SARIMAX (p,d,q) (P,D,Q) m

Seasonal effects
E.g. Temperature

ARIMA(p,d,q)

eXogenous factors
E.g. Humidity
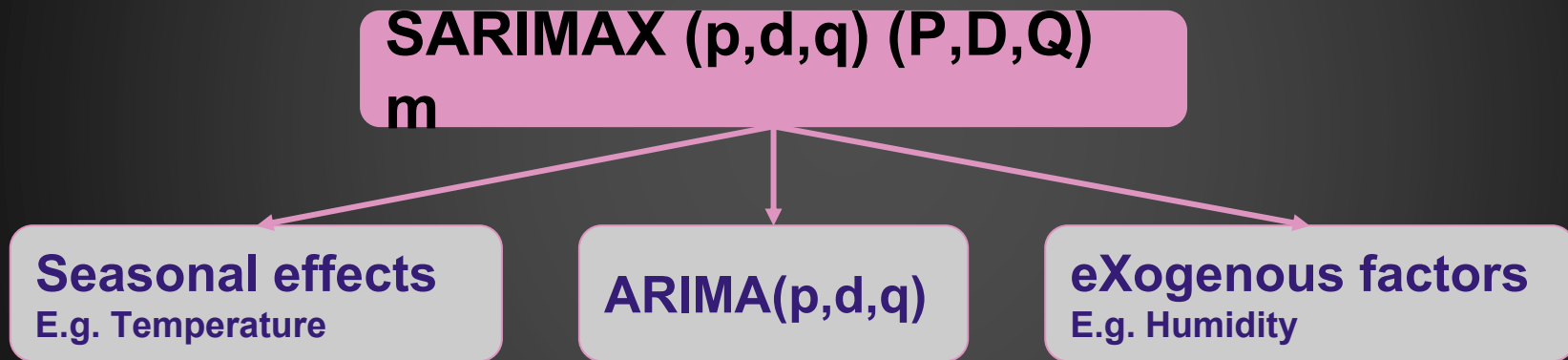
Facebook Prophet

- Forecasting time series data based on an additive regressive model
- Decomposes time series data into trend, seasonality and holiday effect

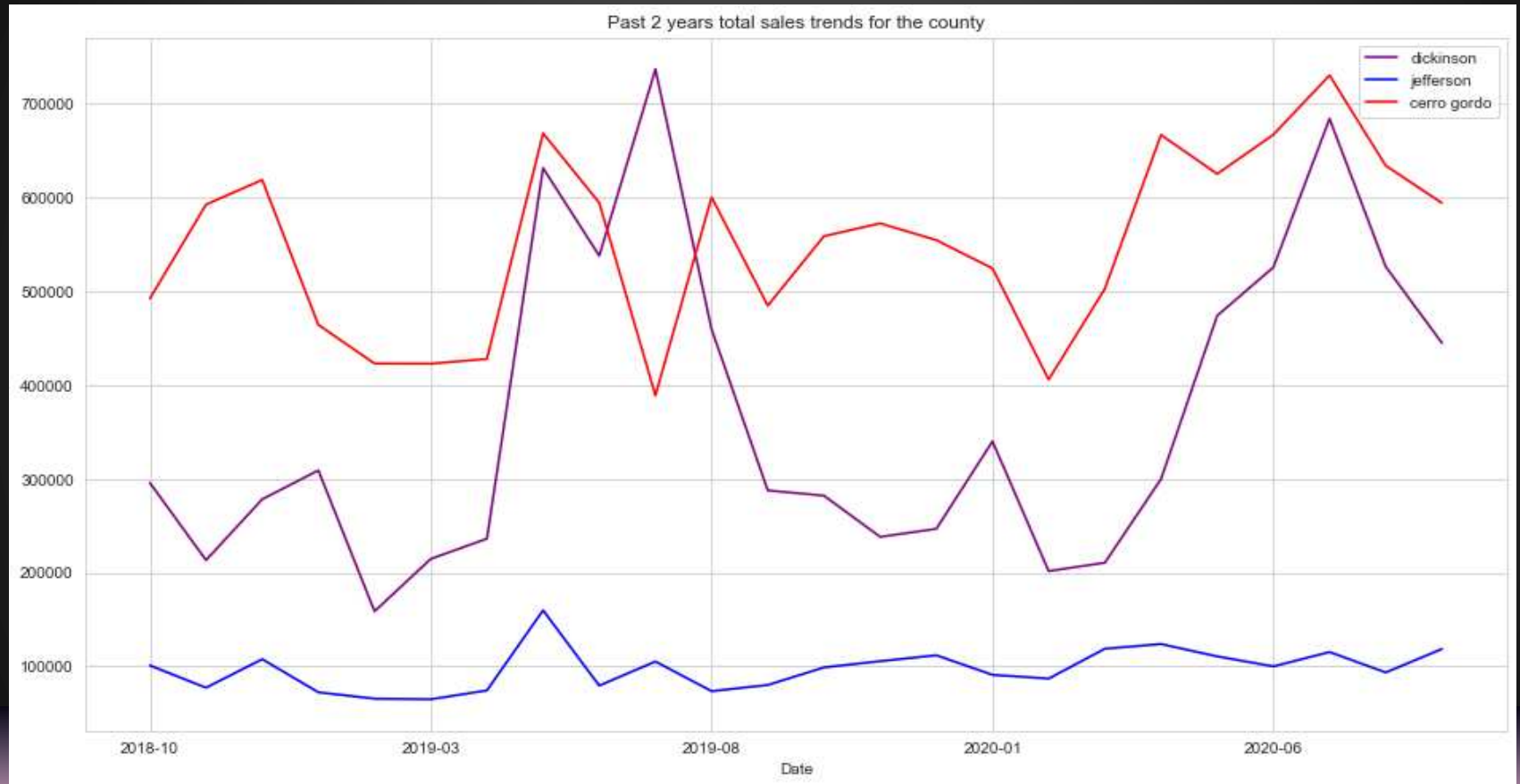$$y(t) = g(t) + s(t) + h(t) + e(t)$$

g(t): a trend function which models the non-periodic changes
s(t): a periodic changes i.e weekly, monthly, yearly.
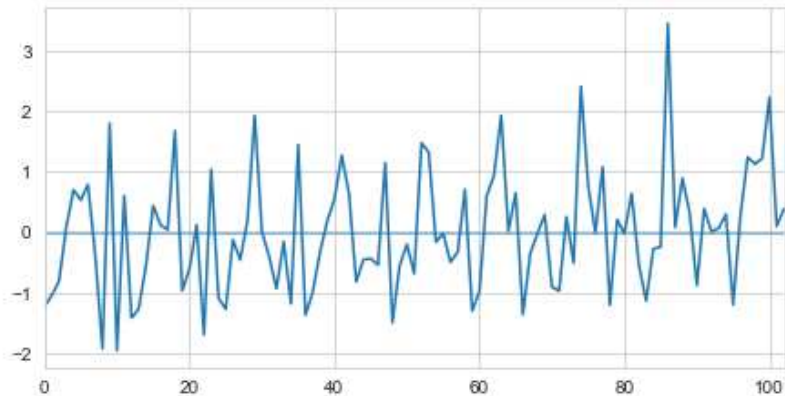**h(t):** effects of holidays (user provided) with irregular schedules
e(t): error changes
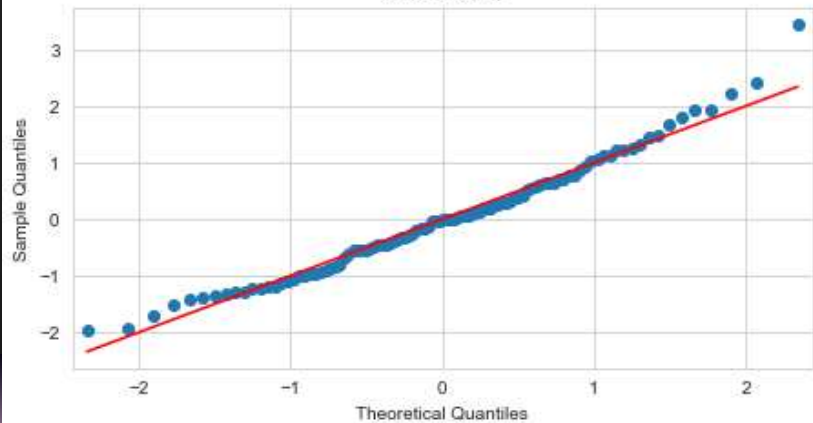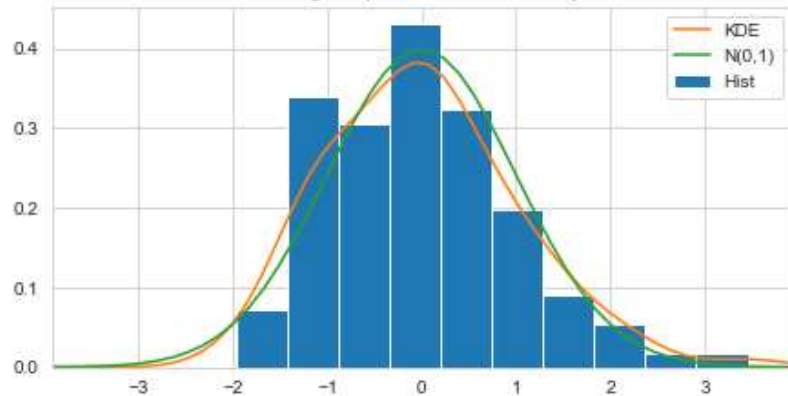
Three Counties Trend

Past 2 years total sales trends for the county
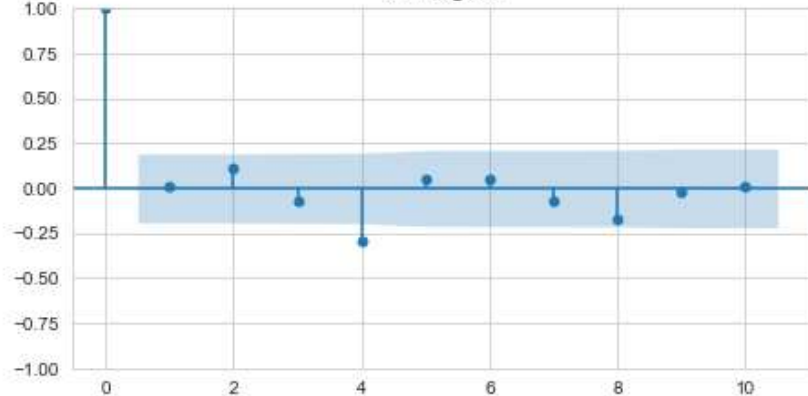
ARIMA

ARIMA Actual vs Forecast

Forecast vs Actuals
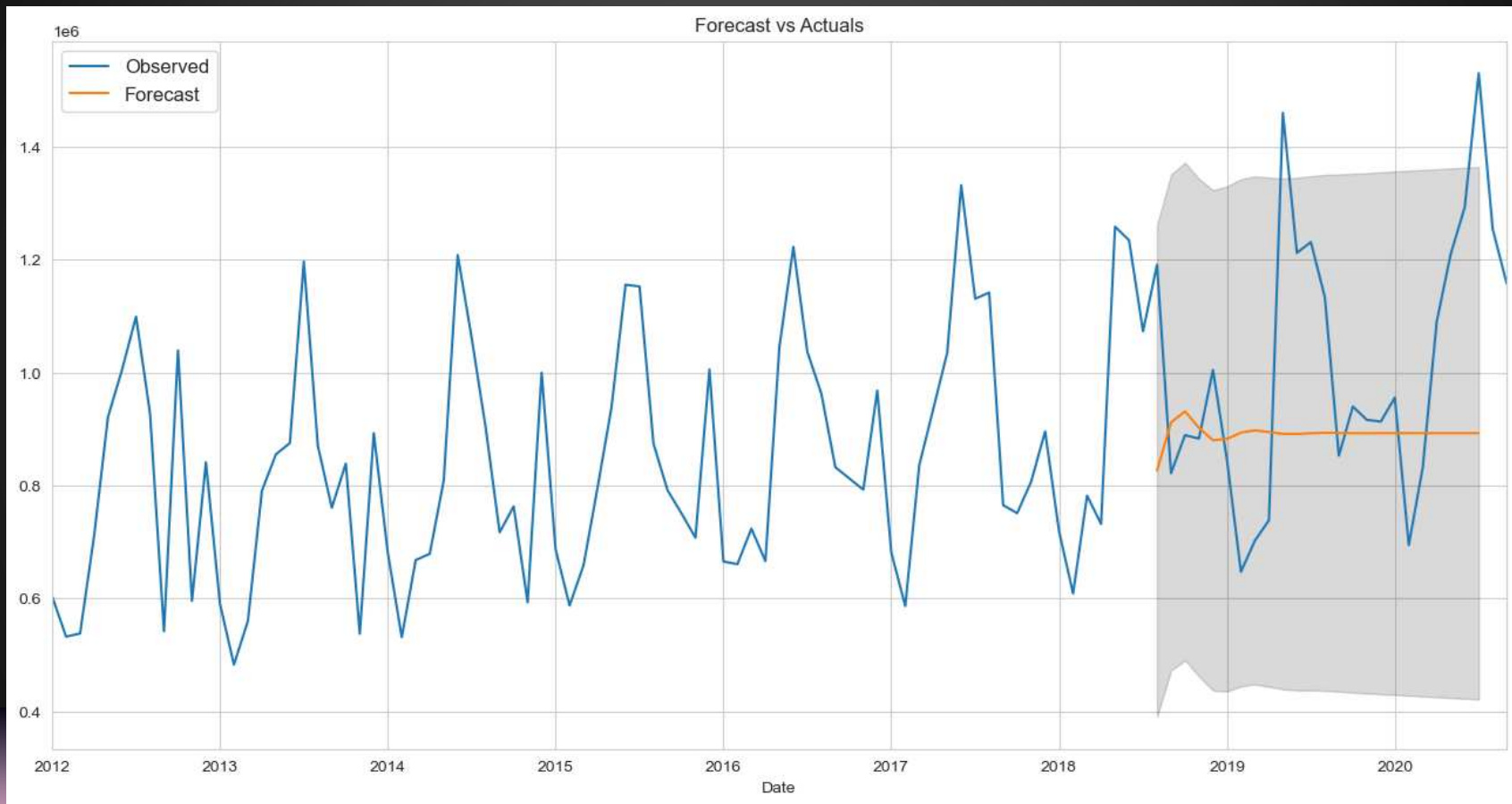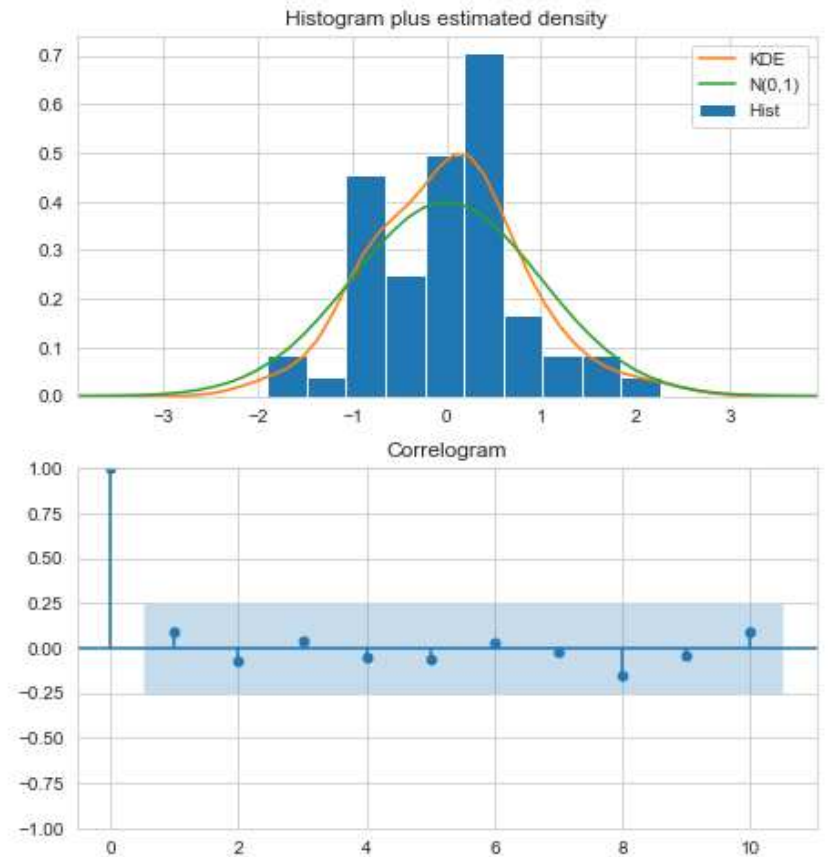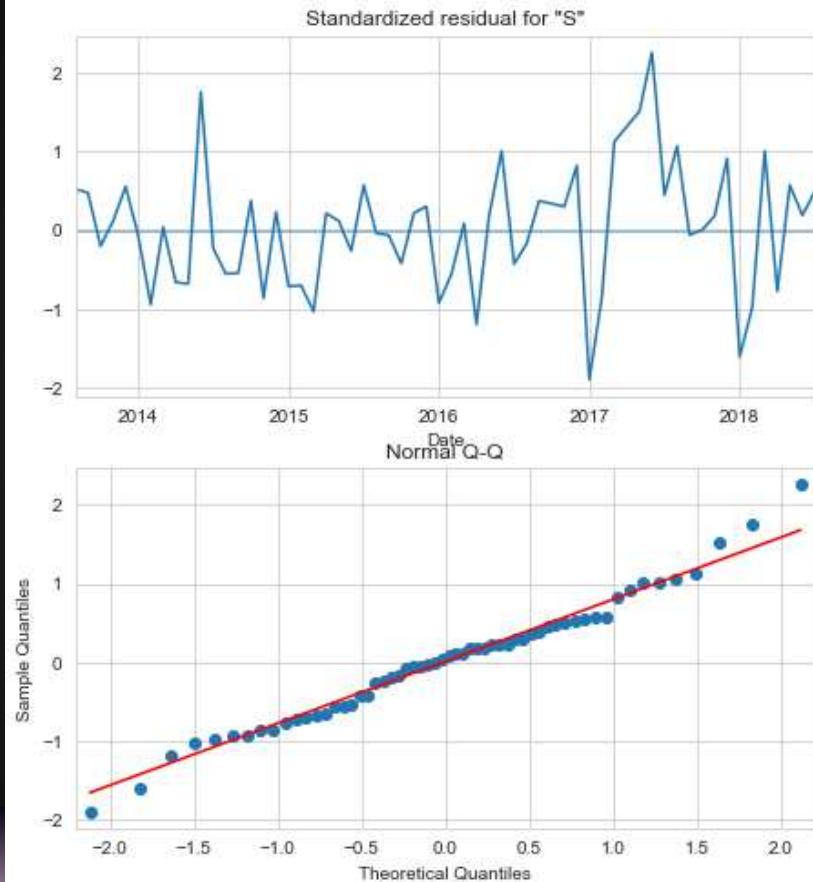
SARIMAX

SARIMAX Actual vs Forecast

Forecast vs Actuals

FB Prophet Model

Total Sales
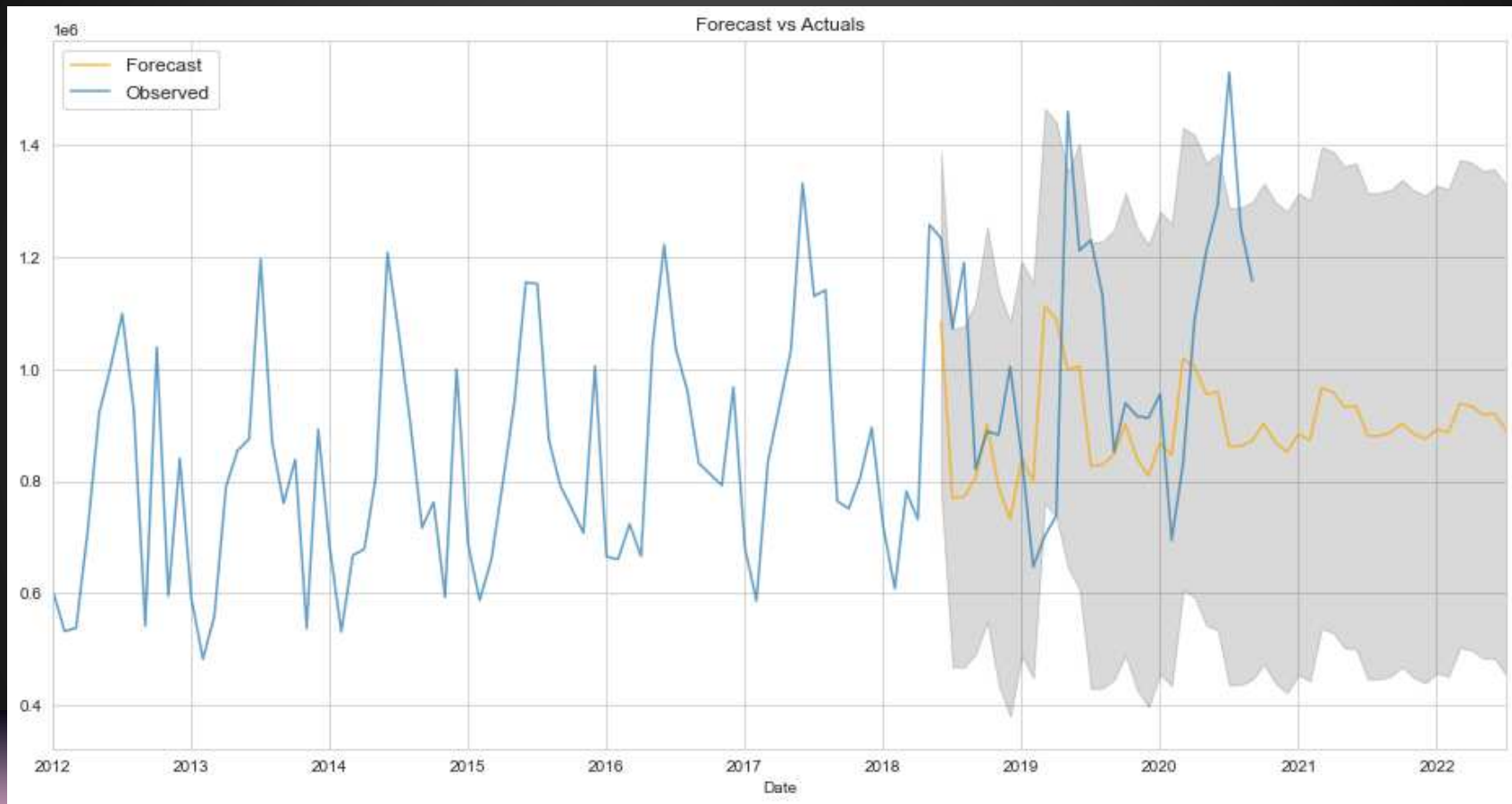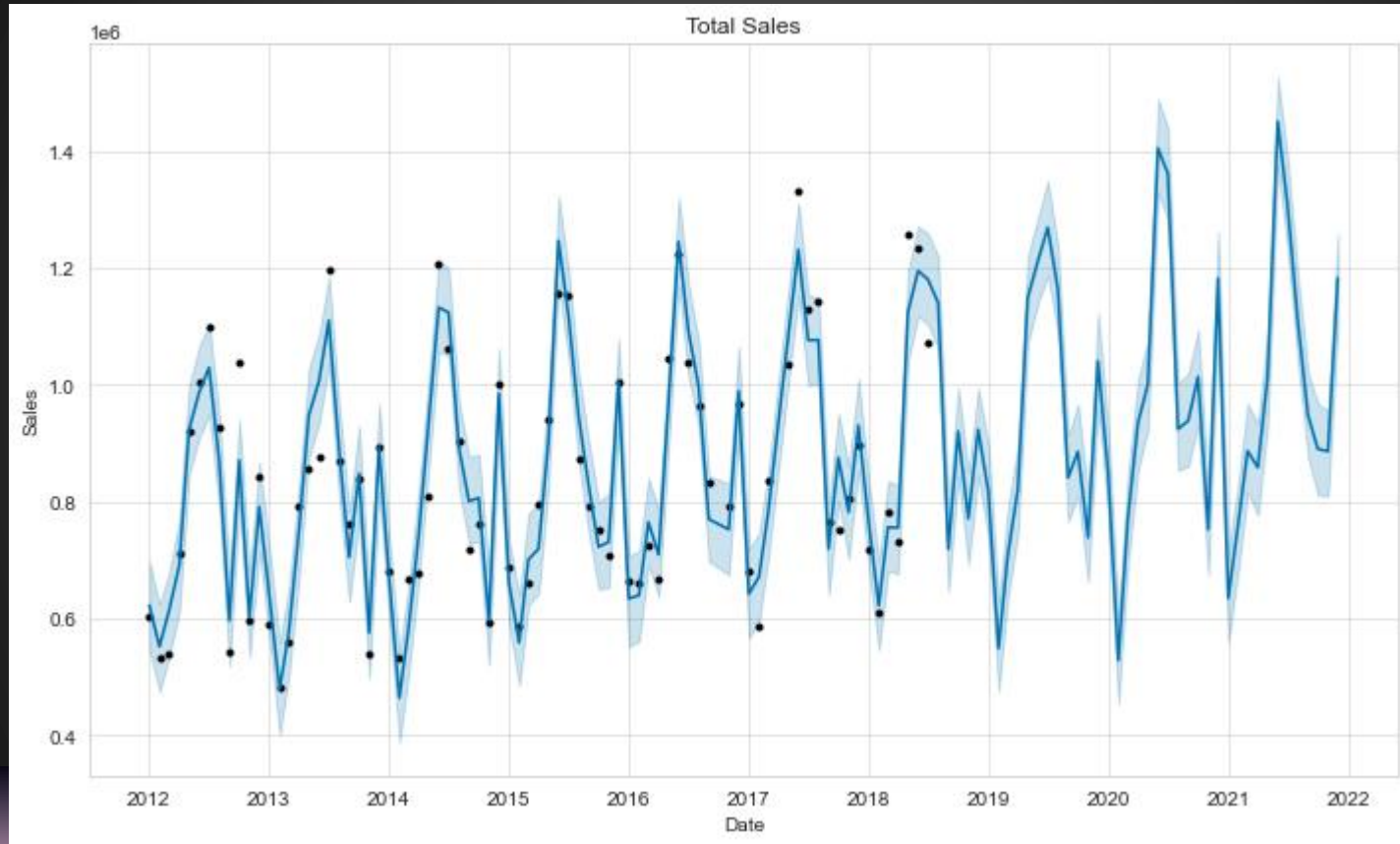
FB Prophet Model

FB Prophet Model

FB Prophet Actual vs Forecast
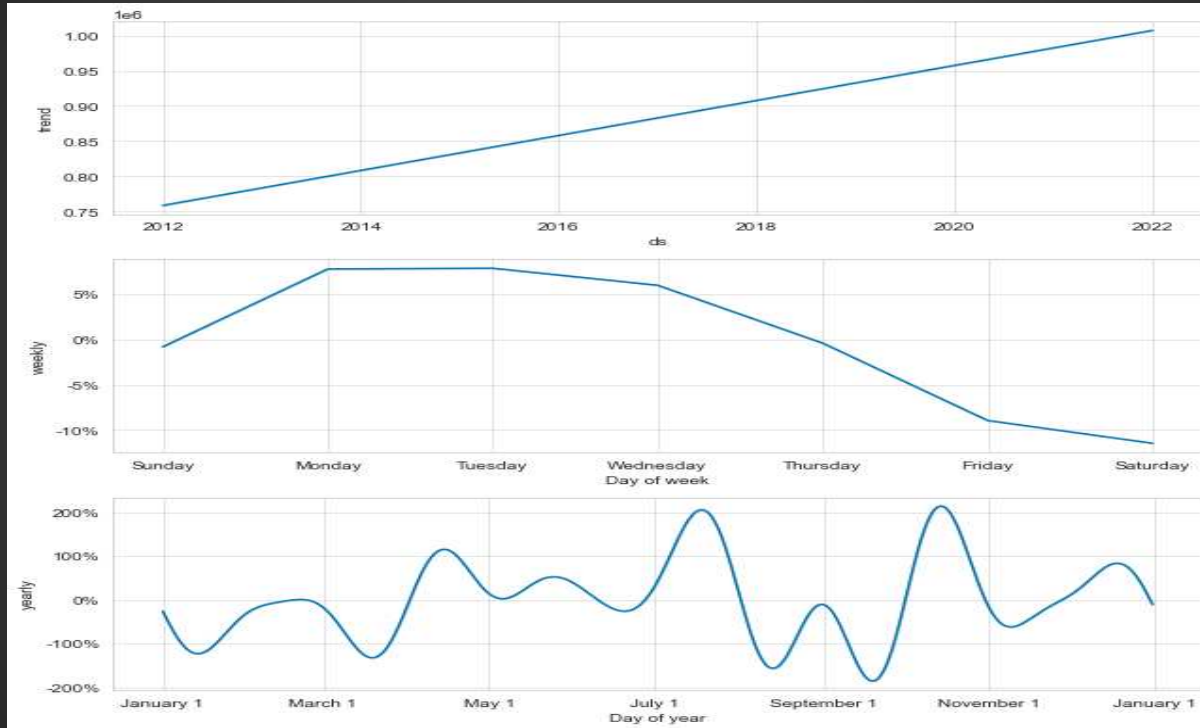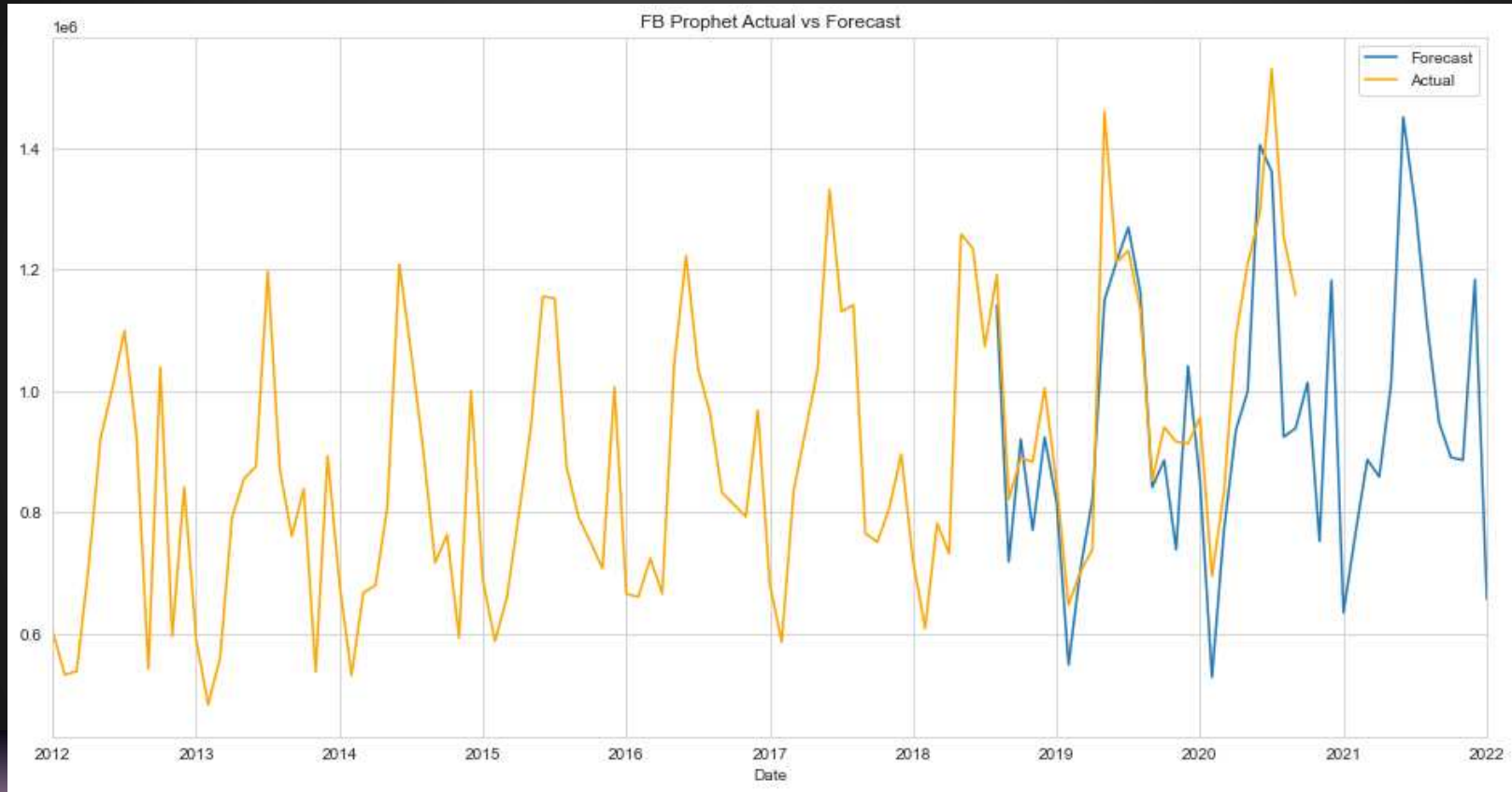
Evaluation Metrics

Model	MSE	RMSE	MAE
ARIMA	54,434,385,279.0	233,311.78	202,573.79
SARIMAX	78,075,356,102.92	279,419.68	221,991.9
FB Prophet	**19,588,716,071.92**	**139,959.69**	**110,473.41**

Dickinson

Dickinson

FB Prophet Actual vs Forecast

Dickinson

41

Cerro Gordo

Cerro Gordo

Cerro Gordo

FB Prophet Actual vs Forecast
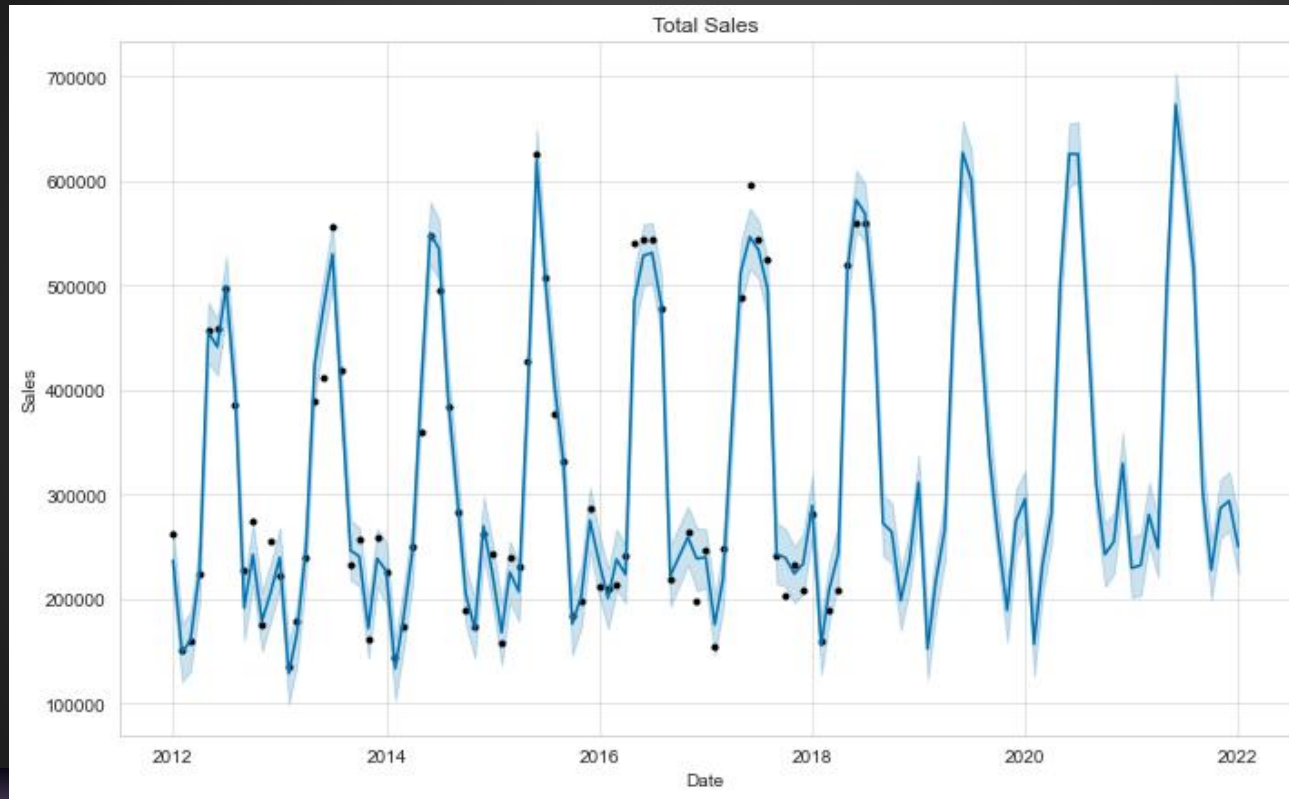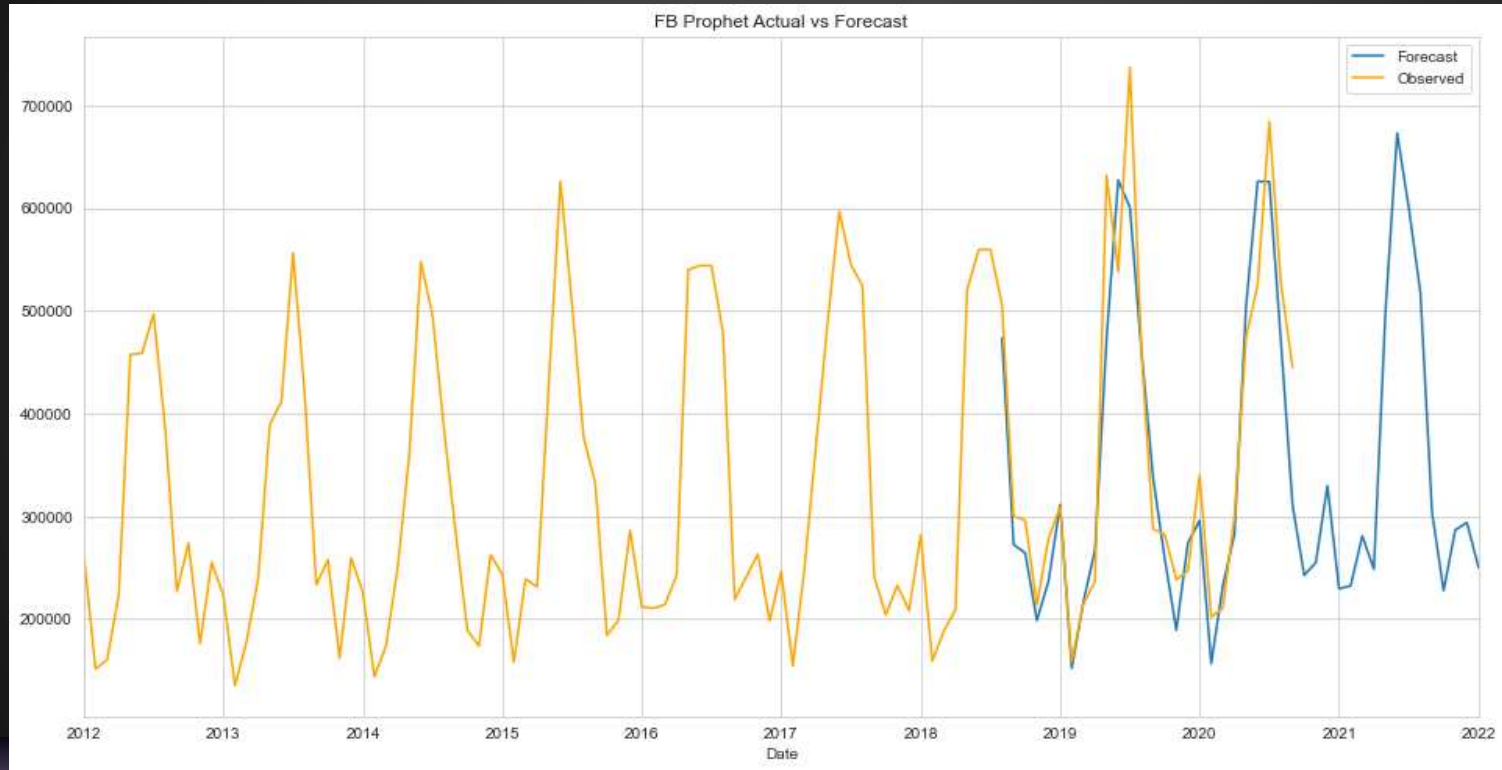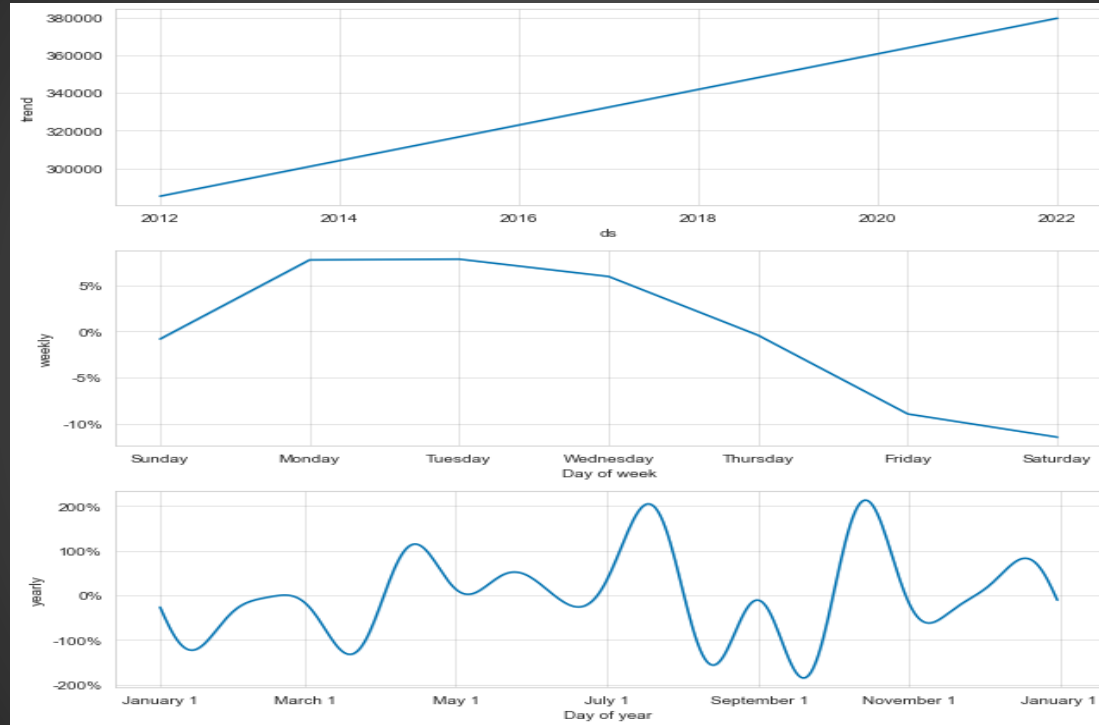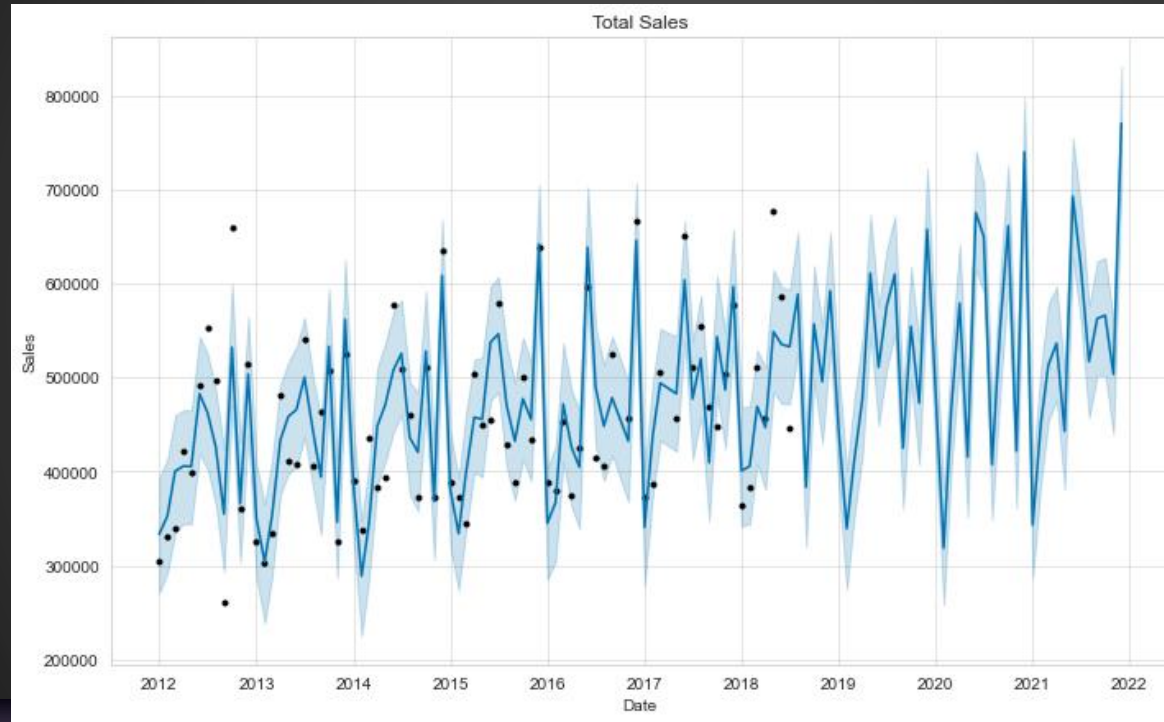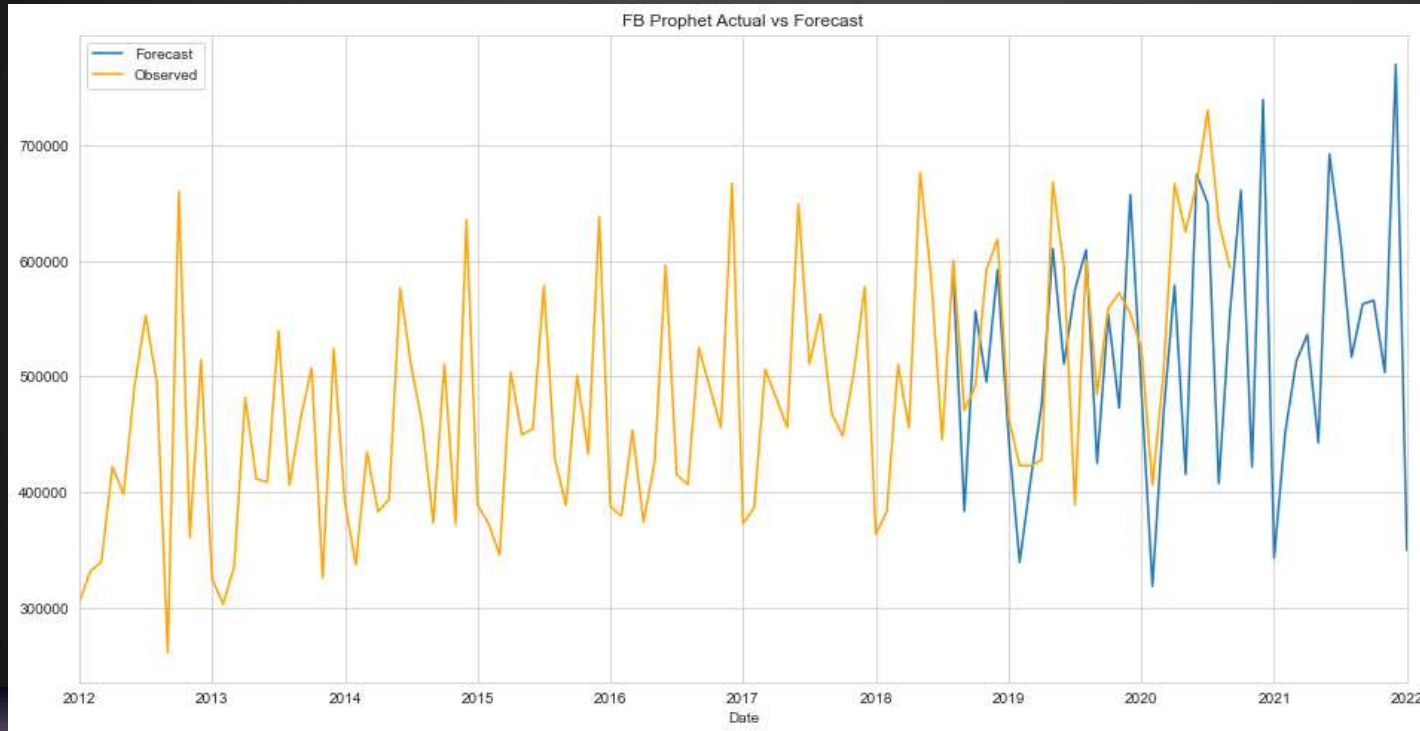
Jefferson

Jefferson

Jefferson

FB Prophet Actual vs Forecast
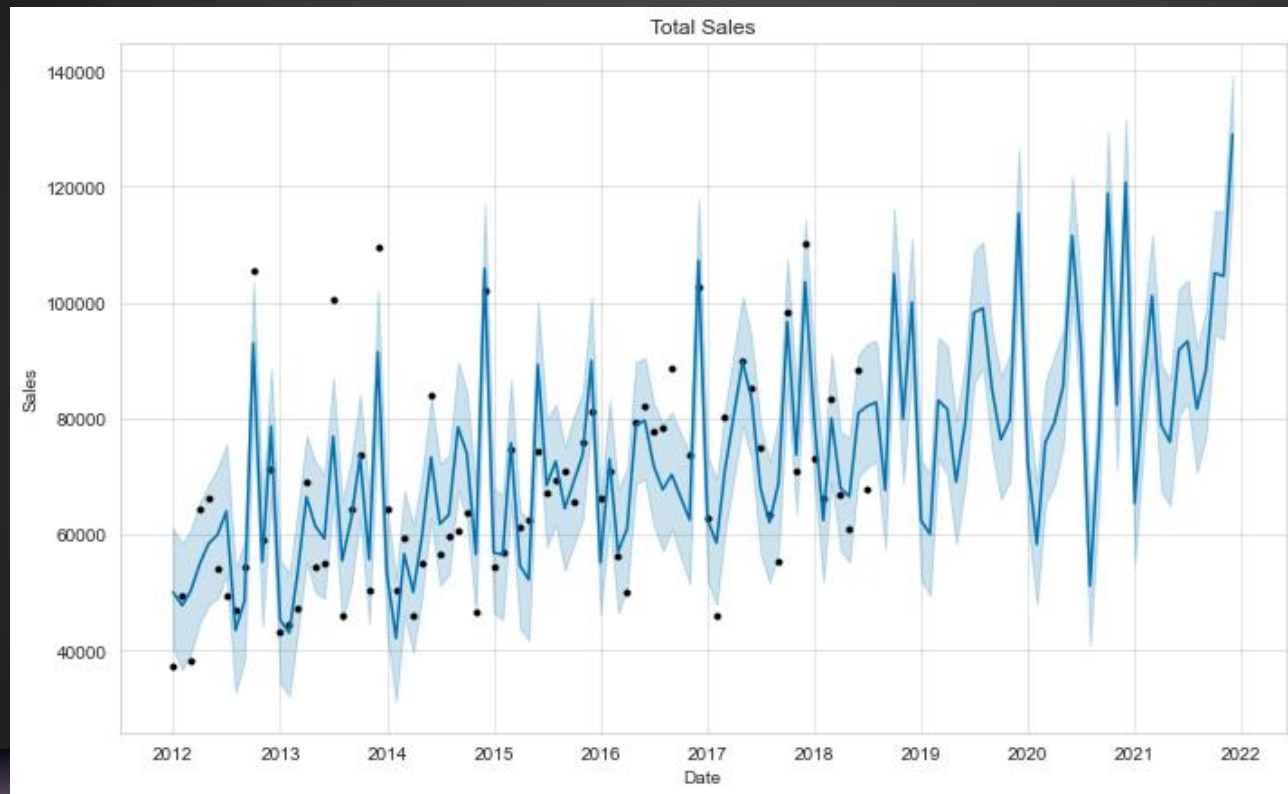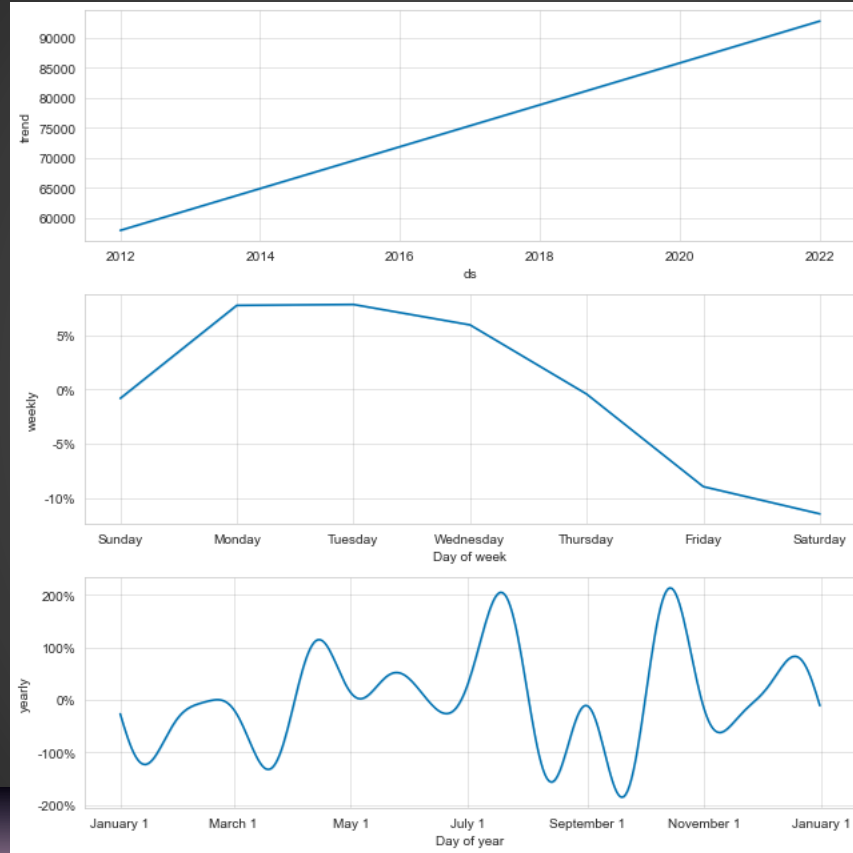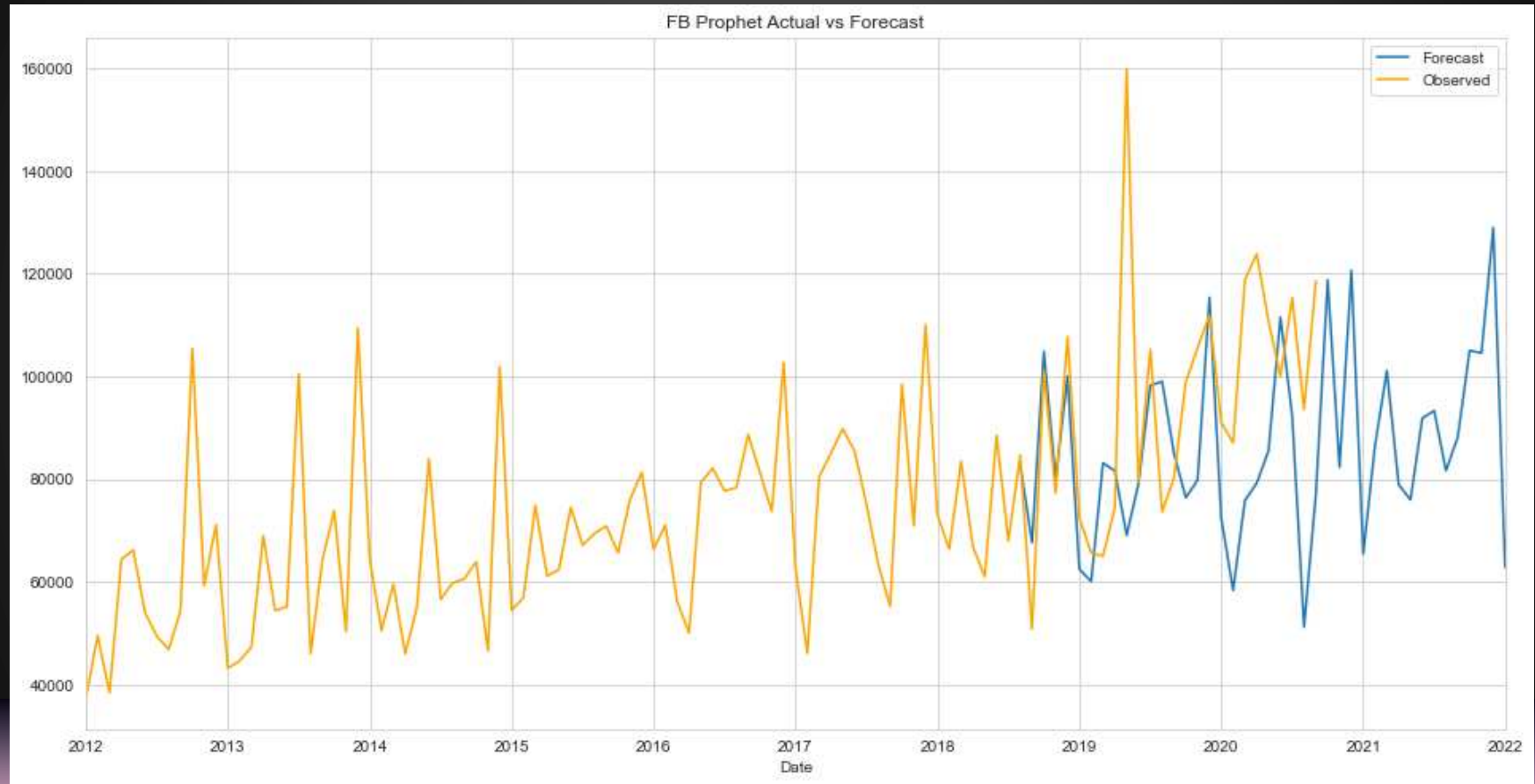
Evaluation Metrics

Model	MSE	RMSE	MAE
Dickinson	3,989,625,541.06	63,163.48	47,379.22
Cerro Gordo	8,562,582,540.07	92,534.22	71,722.83
Jefferson	802,734,699.72	28,332.57	20,488.71

04 Conclusion

Top 10 items we choose to sell in Dickinson

10 items will be sold in Dickinson

1	Captain Morgan Spiced Rum	6	Titos Handmade Vodka
2	Black Velvet	7	Barton Vodka
3	Crown Royal Canadian Whisky	8	Grey Goose Vodka
4	Jack Daniels Old #7 Black Lbl	9	Malibu Coconut Rum
5	Fireball Cinnamon Whiskey	10	Titos Handmade Vodka

Top 10 items we choose to sell in Jefferson/Cerro Gordo

10 items will be sold in Jefferson

1. Black Velvet
2. Hawkeye Vodka
3. Jack Daniels Old #7 Black Lbl
4. Templeton Rye
5. Titos Vodka
6. Captain Morgan Original Spiced
7. Crown Royal
8. Captain Morgan Spiced Rum
9. Fireball Cinnamon Whiskey
10. Jameson

10 items will be sold in Cerro Gordo

1. Black Velvet
2. Captain Morgan Spiced Rum
3. Hawkeye Vodka
4. Fireball Cinnamon Whiskey Mini Dispenser
5. Crown Royal Canadian Whisky
6. Jack Daniels Old #7 Black Lbl
7. Captain Morgan Spiced Barrel
8. Absolut Swedish Vodka 80 Prf
9. Titos Handmade Vodka
10. Black Velvet

Predict Sales

Location	2021 Jan.	2021 Feb.	2021 Mar.	2021 Apr.	2021 May.	2021 Jun.	2021 Jul.	2021 Aug.	2021 Sept.	2021 Oct.	2021 Nov.	2021 Dec.	2022 Jan.
Dickson	229,330	232,196	280,656	248,496	500,644	672,958	600,381	517,784	303,800	227,561	286,686	293,768	250,196
Jefferson	65,411	86,550	101,158	78,909	75,974	91,866	93,331	81,648	88,240	105,080	104,568	129,019	62,824
Cero gordo	343,148	451,502	513,170	536,423	442,848	692,574	618,812	516,957	562,724	566,045	503,472	770,057	349,752

Business Recommendations

Go-in market strategies

01 Go to the market with less competition.

02 Direct acquisition of existing beer stores.

03 Conducting Online Sales.

Product Selling strategies

04 Bundling sale

05 Selected main products.

06 Giveaway promotion.

THANKS!

Do you have any questions?

This presentation template was created by Slidesgo, including icons by **Flaticon**, and infographics & images by **Freepik**

Appendix
Data Quality Report

Overview of data

The 'Iowa liquor sales' data set collects the invoices from 2012/01/03 to 2020/09/30.

Columns includes transactions information, store level info, item level info.

Columns									
1	Invoice/Item Number	6	City	11	Category	16	Item Description	21	Bottles Sold
2	Date	7	Zip Code	12	Category Name	17	Pack	22	Sale (Dollars)
3	Store Number	8	Store Location	13	Vendor Number	18	Bottle Volume (ml)	23	Volume Sold (Liters)
4	Store Name	9	County Number	14	Vendor Name	19	State Bottle Cost	24	Volume Sold (Gallons)
5	Address	10	County	15	Item Number	20	State Bottle Retail		

Statistic summary of numerical columns

	Pack	Bottle Volume (ml)	State Bottle Cost	State Bottle Retail	Bottles Sold	Sale (Dollars)	Volume Sold (Liters)
mean	12.3	911.7	9.6	14.7	8.8	128.2	8.9
std	7.6	604.2	9.8	14.8	12.8	456.1	33.0
min	-96.0	0.0	0.0	0.0	-128.0	0.0	0.0
25%	6.0	750.0	5.5	8.3	2.0	31.5	1.5
50%	12.0	750.0	8.2	12.3	6.0	71.5	5.0
75%	12.0	1000.0	12.1	18.1	12.0	139.3	10.5
max	120.0	378000.0	6468.0	9702.0	127.0	279557.3	15000.0

Completeness of data

Most columns has no missing value, except store location.

Columns name	Number of Missing Value	% of missing value
Invoice/Item Number	0	0%
Date	0	0%
Store Number	0	0%
Store Name	0	0%
Address	79927	0%
City	79926	0%
Zip Code	79971	0%
Store Location	1886023	10%

Completeness of data

Columns name	Number of Missing Value	% of missing value
County Number	156731	1%
County	156729	1%
Category	16974	0%
Category Name	25040	0%
Vendor Number	5	0%
Vendor Name	3	0%
Item Number	0	0%
Item Description	0	0%

Completeness of data

Columns name	Number of Missing Value	% of missing value
Pack	0	0%
Bottle Volume (ml)	0	0%
State Bottle Cost	10	0%
State Bottle Retail	10	0%
Bottles Sold	0	0%
Sale (Dollars)	10	0%
Volume Sold (Liters)	0	0%
Volume Sold (Gallons)	0	0%

Basic information

Number of ...

County	99
Stores	2484
Category	110

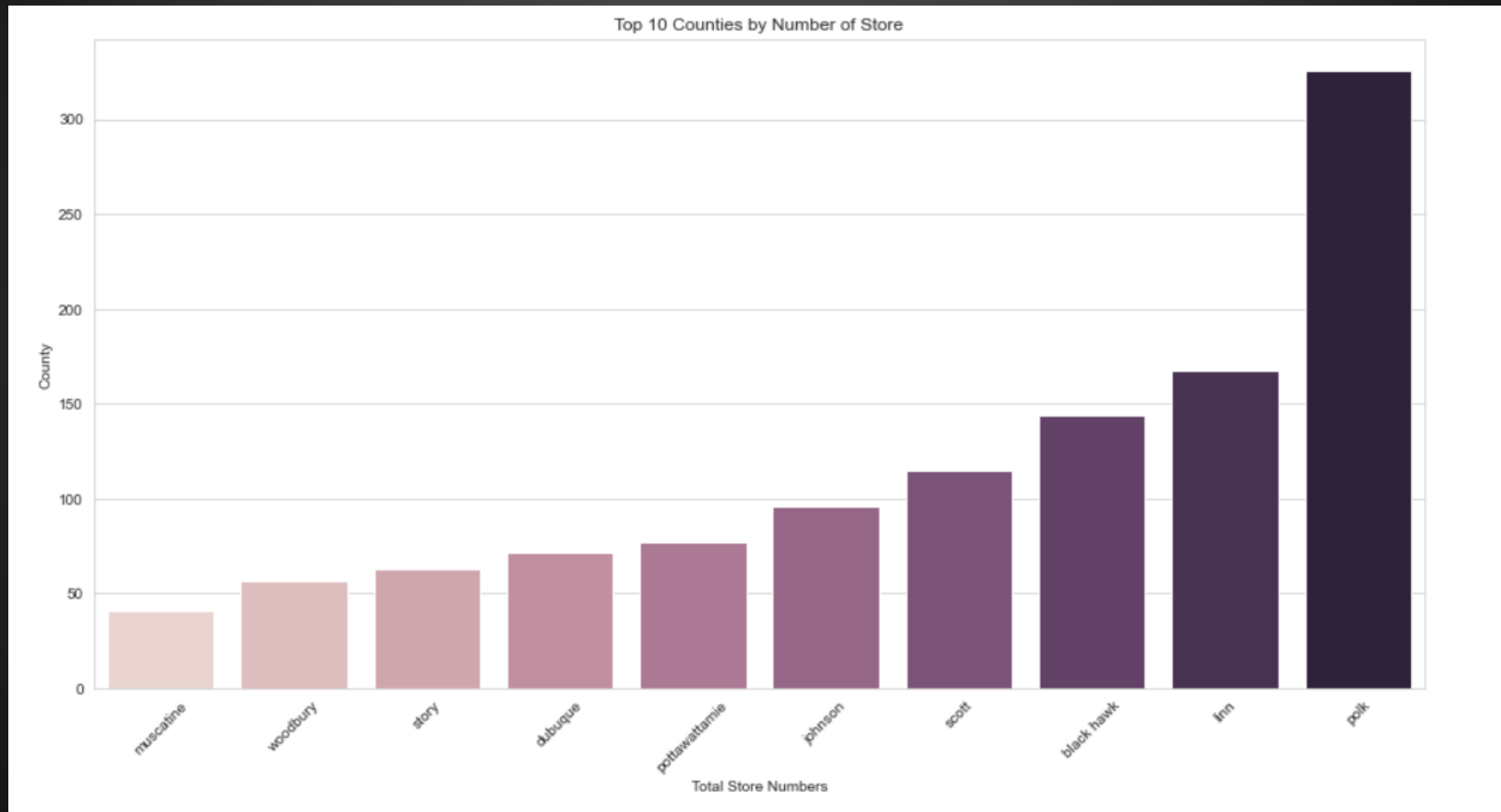|	
Vendor	348
Item	11709

Uniqueness

No duplicated data

Accuracy
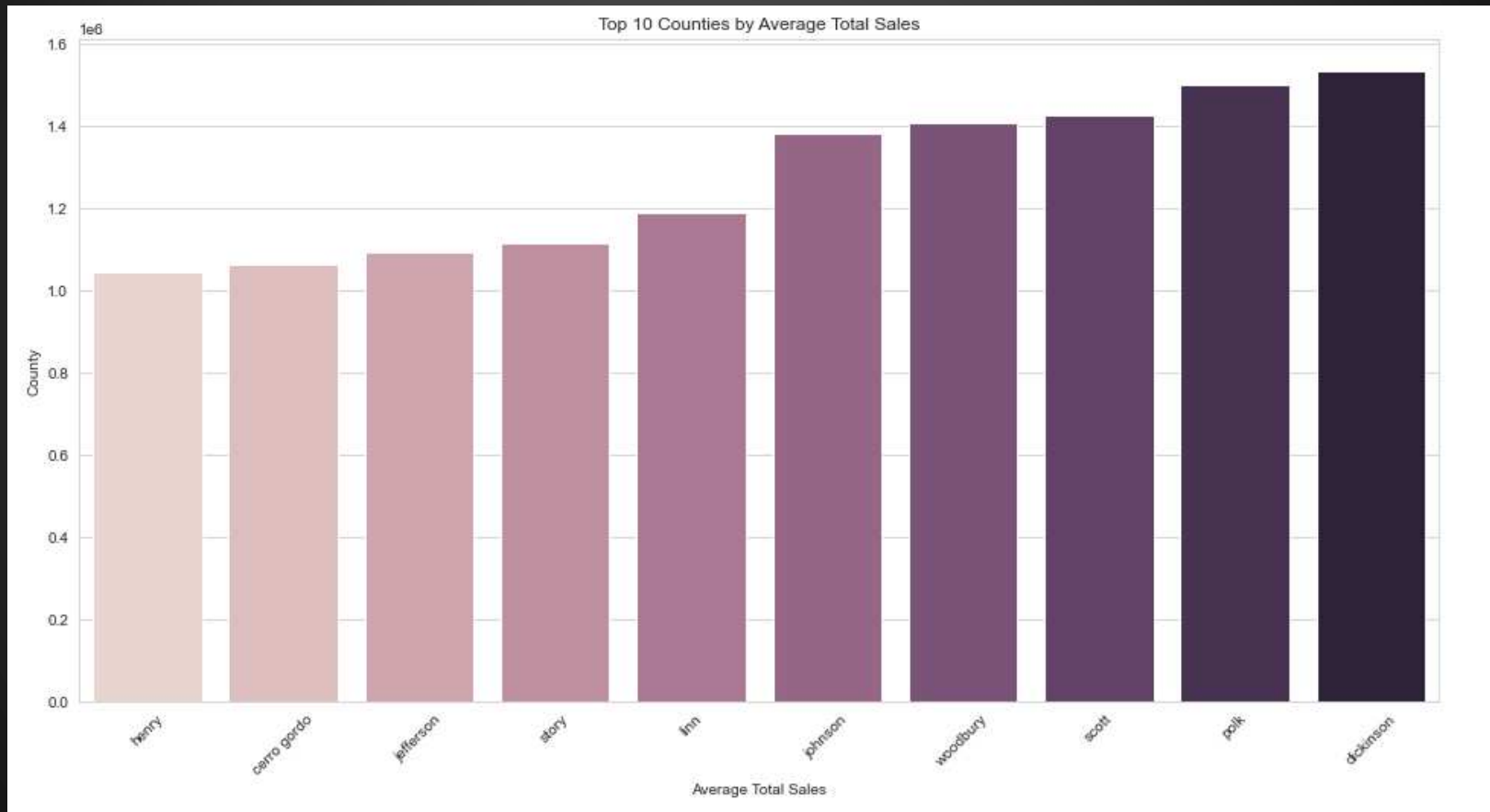
Meaningless text in Store Name, with some punctuations and numbers.

Different Item Number but same Item description

Negative values in Pack & Bottles sold

Number of stores information

Top 10 counties by Store Average Total Sales

Top 10 Counties by Average Total Sales

END

This presentation template was created by Slidesgo, including icons by **Flaticon**, and infographics & images by **Freepik**