Weekly Summary - Group 5

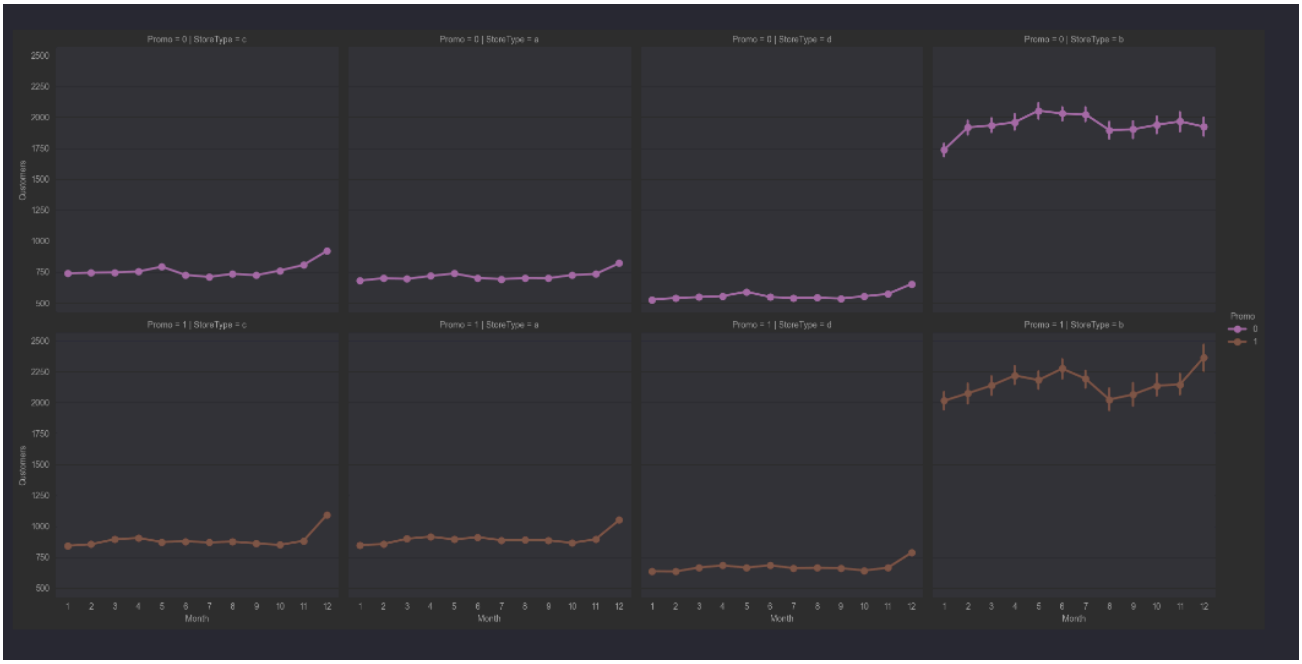
Emily Ye zy367@drexel.edu
Junkai Ge jg3944@drexel.edu
Jerry Li jl4533@drexel.edu
Shengyang Dong sd3666@drexel.edu

Summary

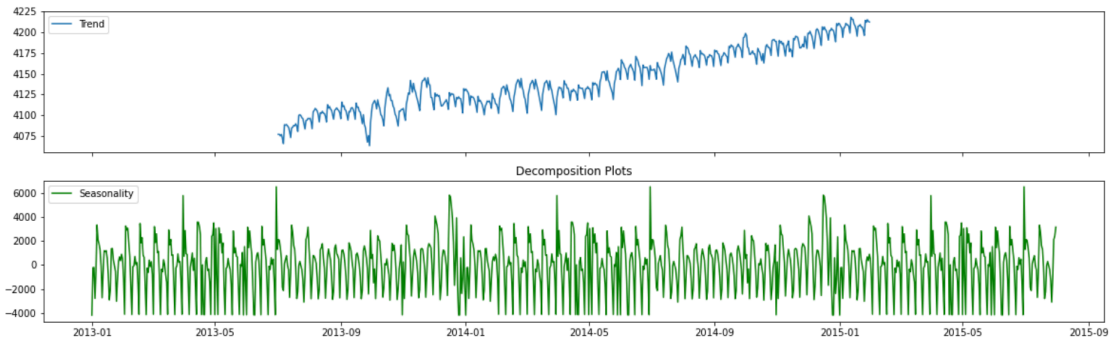
Thank you for your feedback on our last submission. Following your advice, we have added more exploratory data analysis (Part 1) to gain better insights into the dataset. We have also started building a predictive model using ARIMA for initial analysis. Additionally, we are exploring the Seasonal ARIMA (SARIMA) model to handle seasonality and improve our forecasts.

Shengyang Dong:

- **EDA Analysis**
  - Analyzing store types and their performance in terms of sales and customer numbers.
  - Investigating how promotions and competition affect store types.



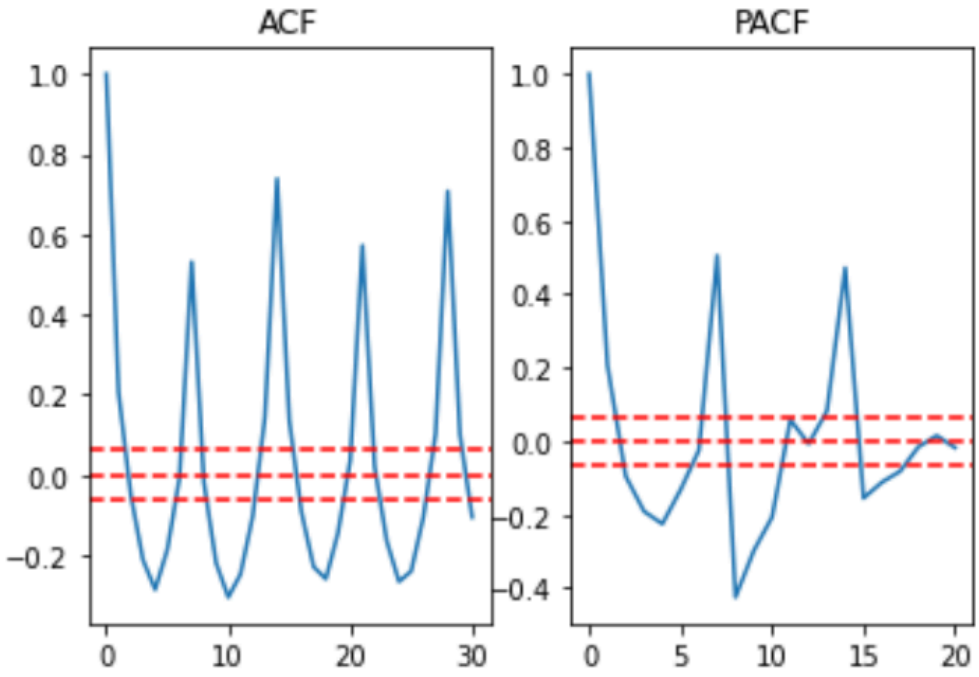
- **Seasonality and Trend Analysis**
  - Implementing code for decomposing sales data into trend, seasonal, and residual components.



Junkai Ge:

- **Seasonality and Trend Analysis**
  - Visualizing and quantifying underlying patterns in the data.
- **Autocorrelation Analysis**

- Utilizing ACF and PACF functions to determine lags for the ARIMA model.
- Visualizing autocorrelation and partial autocorrelation functions.
- Determining ARIMA model parameters (p and q values).



Jerry Li:

- **ARIMA Model Building**
  - Conducting a grid search for optimal ARIMA parameters.
  - Fitting the ARIMA model with the selected parameters.
  - Planning the next steps, including forecasting and validation.

```python
# Define the p, d and q parameters to take any value between 0 and 3
p = d = q = range(0, 2)

# Generate all different combinations of p, q and q triplets
pdq = list(itertools.product(p, d, q))

# Generate all different combinations of seasonal p, q and q triplets
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]

print('Examples of parameter combinations for Seasonal ARIMA: ')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))
```

Emily Ye:

- **SARIMA Model Development**
  - Conducting seasonality analysis and data normalization.
  - Determining SARIMA parameters through spectral analysis and stationarity checks.
  - Fitting the SARIMA model using both Auto ARIMA and manual parameter selection.

SARIMAX Results

Dep. Variable:	Sales	No. Observations:	1017209
Model:	SARIMAX(1, 0, 0)	Log Likelihood	1083036.258
Date:	Fri, 09 Aug 2024	AIC	-2166068.516
Time:	02:43:58	BIC	-2166044.851
Sample:	0	HQIC	-2166062.008
	- 1017209		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.8663	0.000	2470.527	0.000	0.866	0.867
sigma2	0.0070	5.13e-06	1356.503	0.000	0.007	0.007

Ljung-Box (L1) (Q):	217611.75	Jarque-Bera (JB):	1326343.46
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.93	Skew:	0.45
Prob(H) (two-sided):	0.00	Kurtosis:	8.52

Future Work

In our future work, we plan to enhance the predictive accuracy of our model by aggregating sales data at the store level, which will help capture store-specific trends and patterns. Additionally, we will incorporate holidays as exogenous variables by explicitly modeling them with a binary indicator, allowing the model to account for the impact of holidays on sales. Lastly, we will focus on fine-tuning the model to further improve its performance and reliability.