

LET'S TALK ABOUT

FORECASTING COMMERCIAL BANKS TRANSACTIONS

BY

-
1. NOBERT AKWIR
 2. KAREN AMANYA
 3. EUGENE KULOBA
 4. DAVID MWITI
 5. ANTHONY NENE
-

BUSINESS UNDERSTANDING

NCBA Bank in Kenya is experiencing challenges in managing cash levels at its branches. Despite increased access to banking services, many customers still prefer physical bank branches. This has led to increased requests for cash from the central vault, and sometimes branches breach their cash limits. NCBA Bank aims to develop a time series analysis model to forecast future transaction thresholds to help branches manage cash levels more effectively and avoid over-limit breaches.

OBJECTIVES

General Objective

The general objective is to develop a model that will successfully forecast cash transactions that will help NCBA Bank maintain a sufficient cash at its specific branches

Specific Objectives

- Help the bank to sufficiently meet customers' withdrawal needs without fail.
- Help the bank in effectively cutting down on costs related to having excess cash at a branch i.e vault storage services, insurance, vault-to-branch transfer service costs
- Help the bank reduce the risk and adverse effects of robbery

DATA UNDERSTANDING



The data used in this project is sourced from data world. The data consists of demo bank transaction details (debits and credits) carried out by different customers and includes other details such as the amount, the mode of transaction (either cash, RTGS, EFT, SWIFT e.t.c), the account's balance at the time of the transaction, the time of the transaction e.t.c

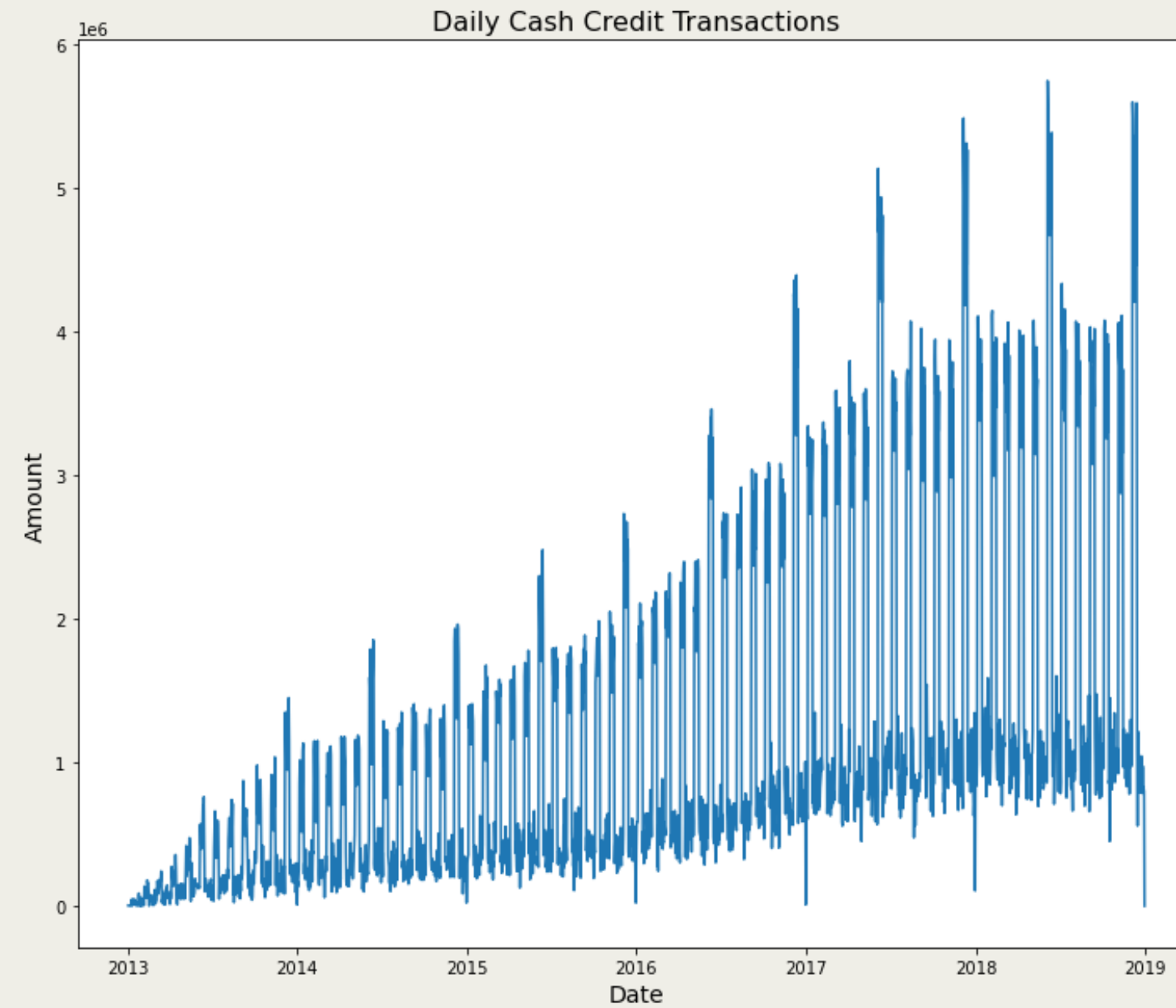


The dataset contains 1,056,320 rows with 16 columns. Relevant columns based on our objectives were:

- Type - describe whether a transaction is a credit or a debit.
- Operation - has a narration of the particular type of transaction.
- Amount - transaction amount.
- full date - date of the transaction.

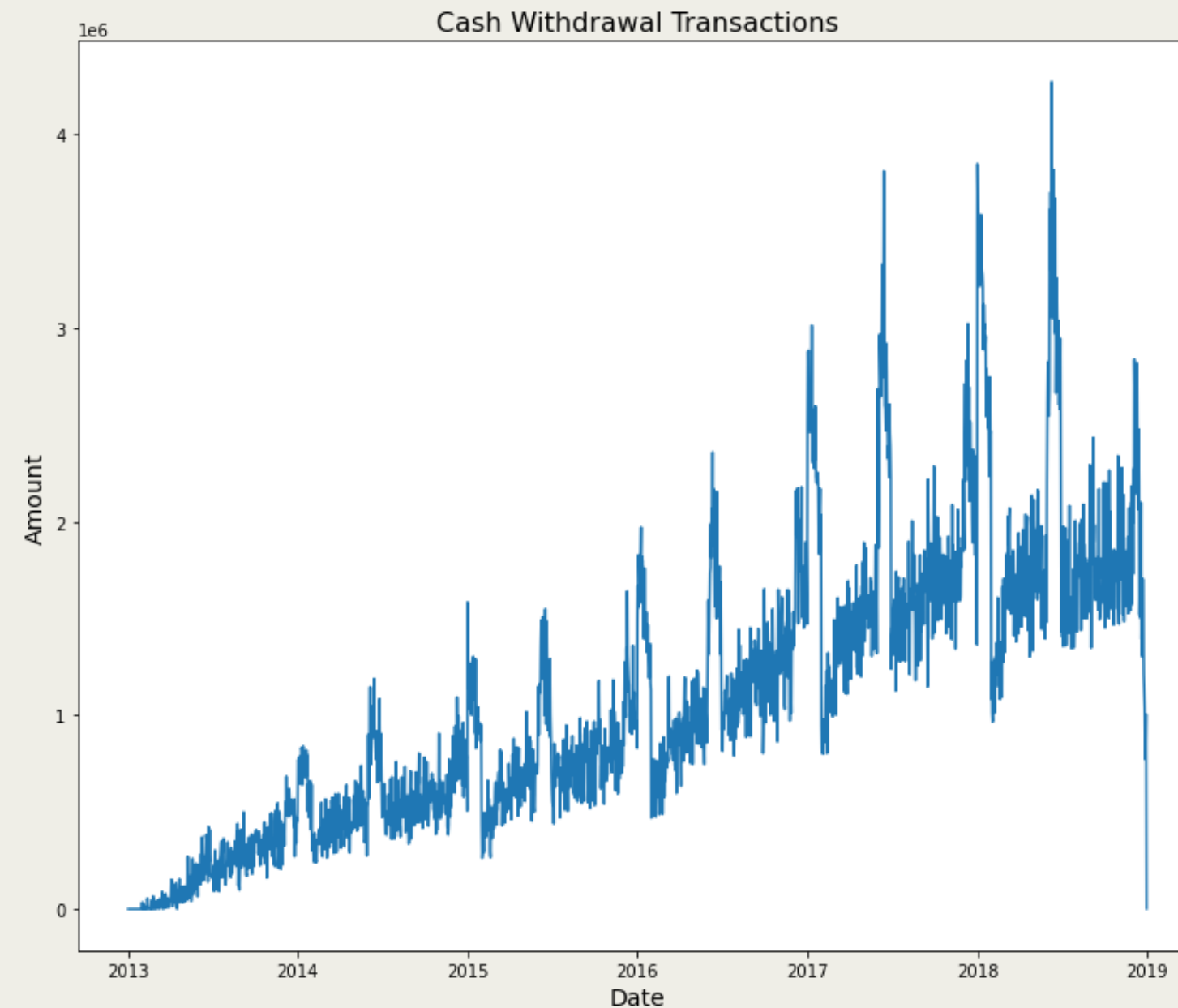


- All Debit transactions were converted to negative entries
- Operations not related to hard cash were removed
- Separate columns for withdrawals and credits were created.
- Time series was downsampled to days



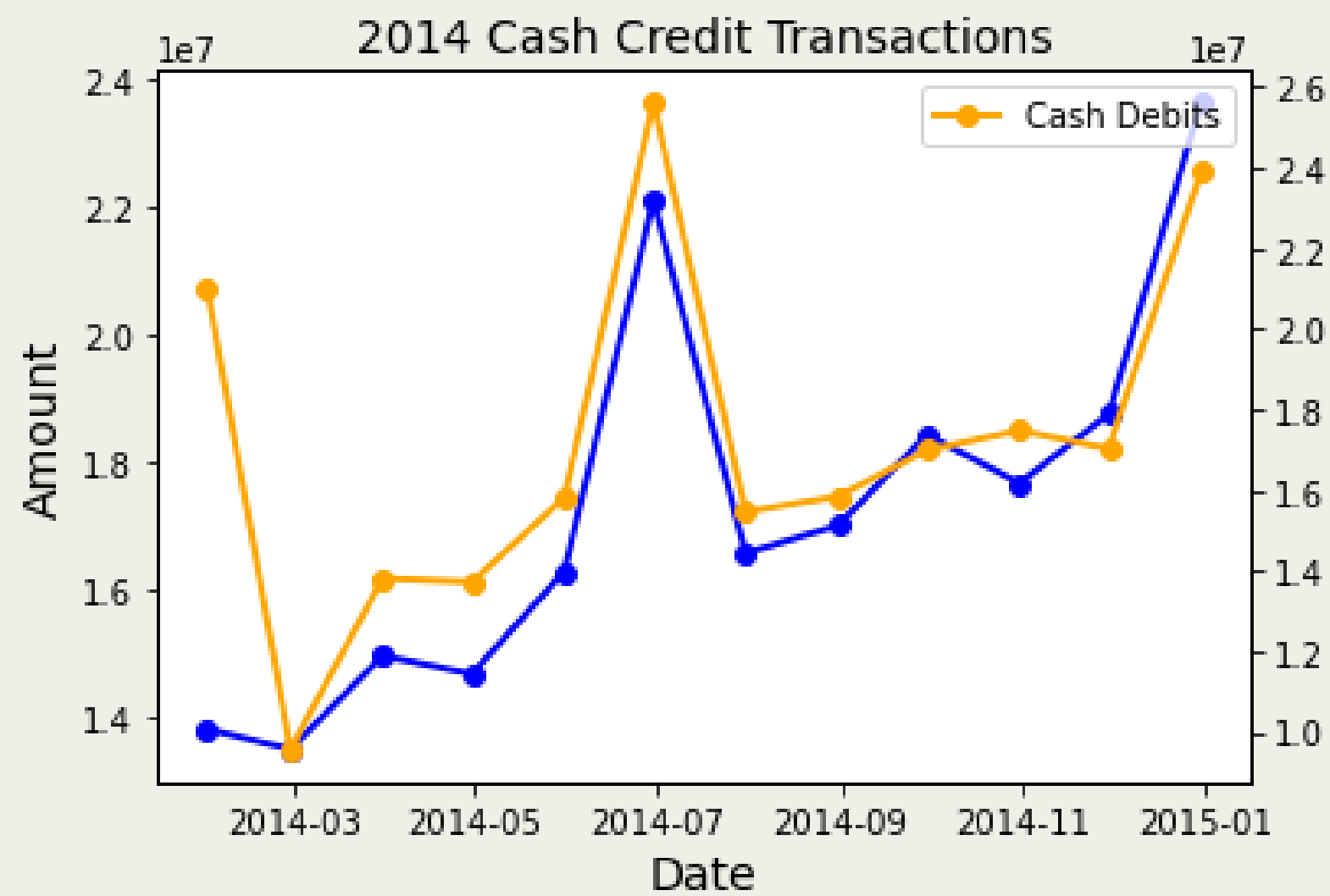
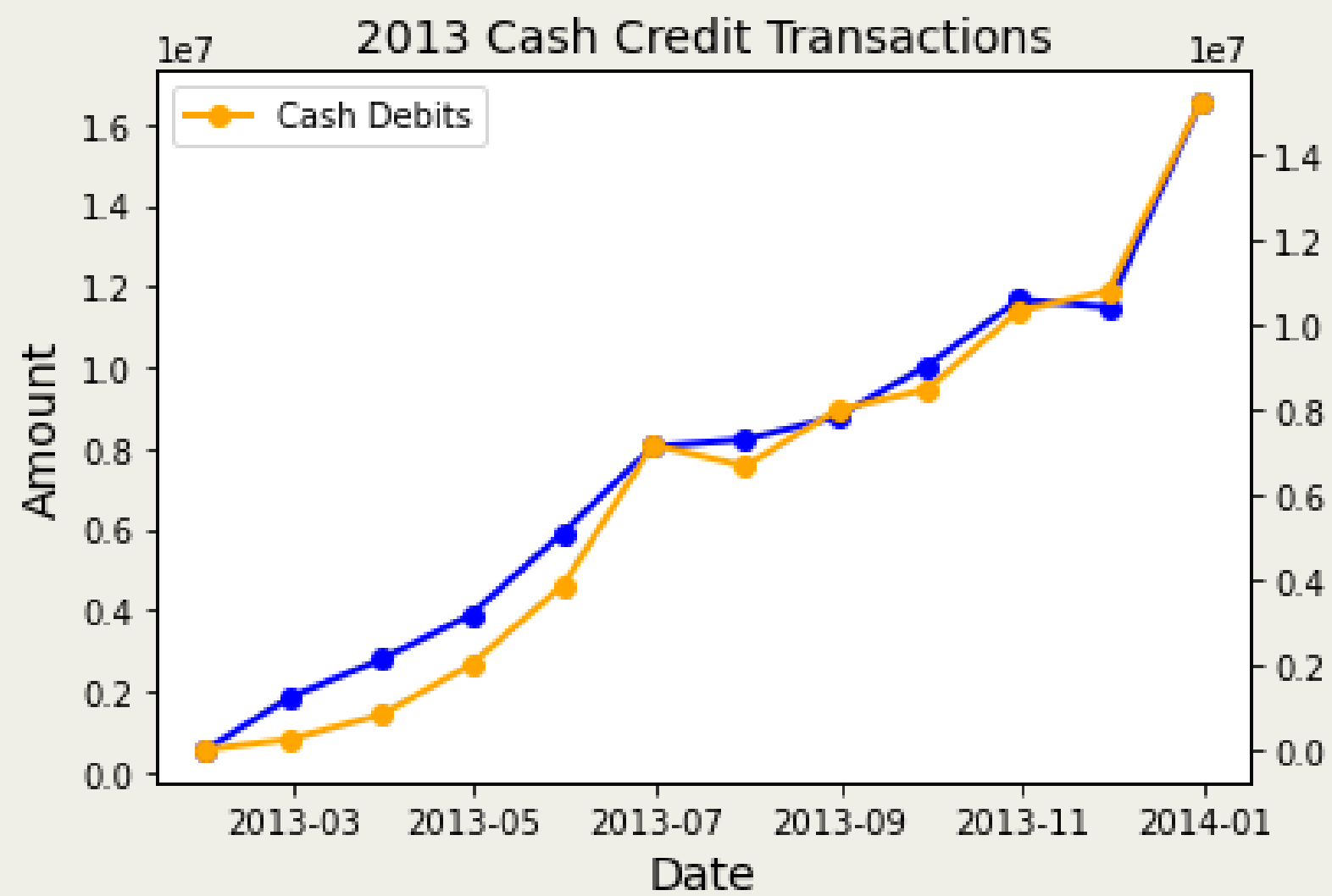
Observations:

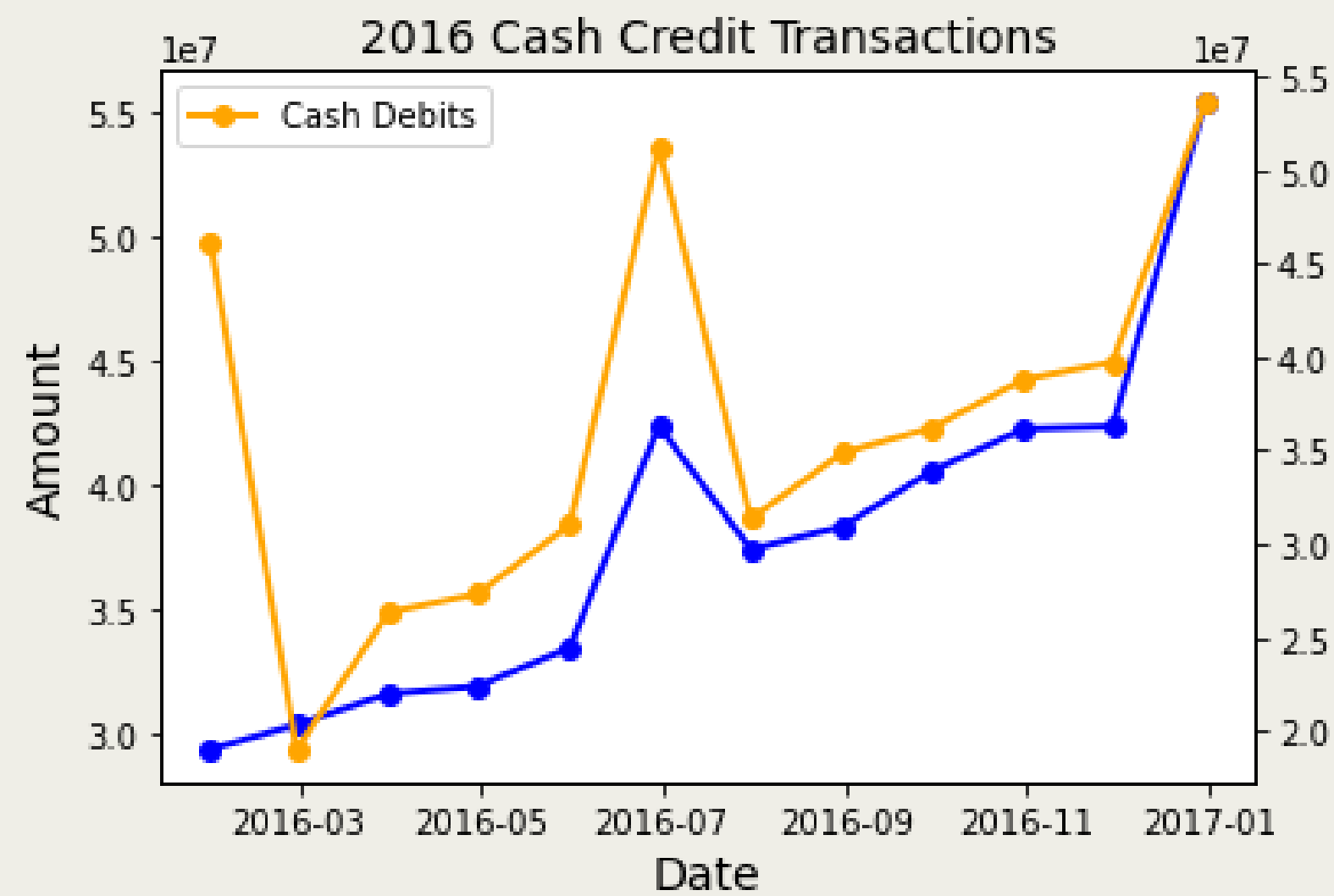
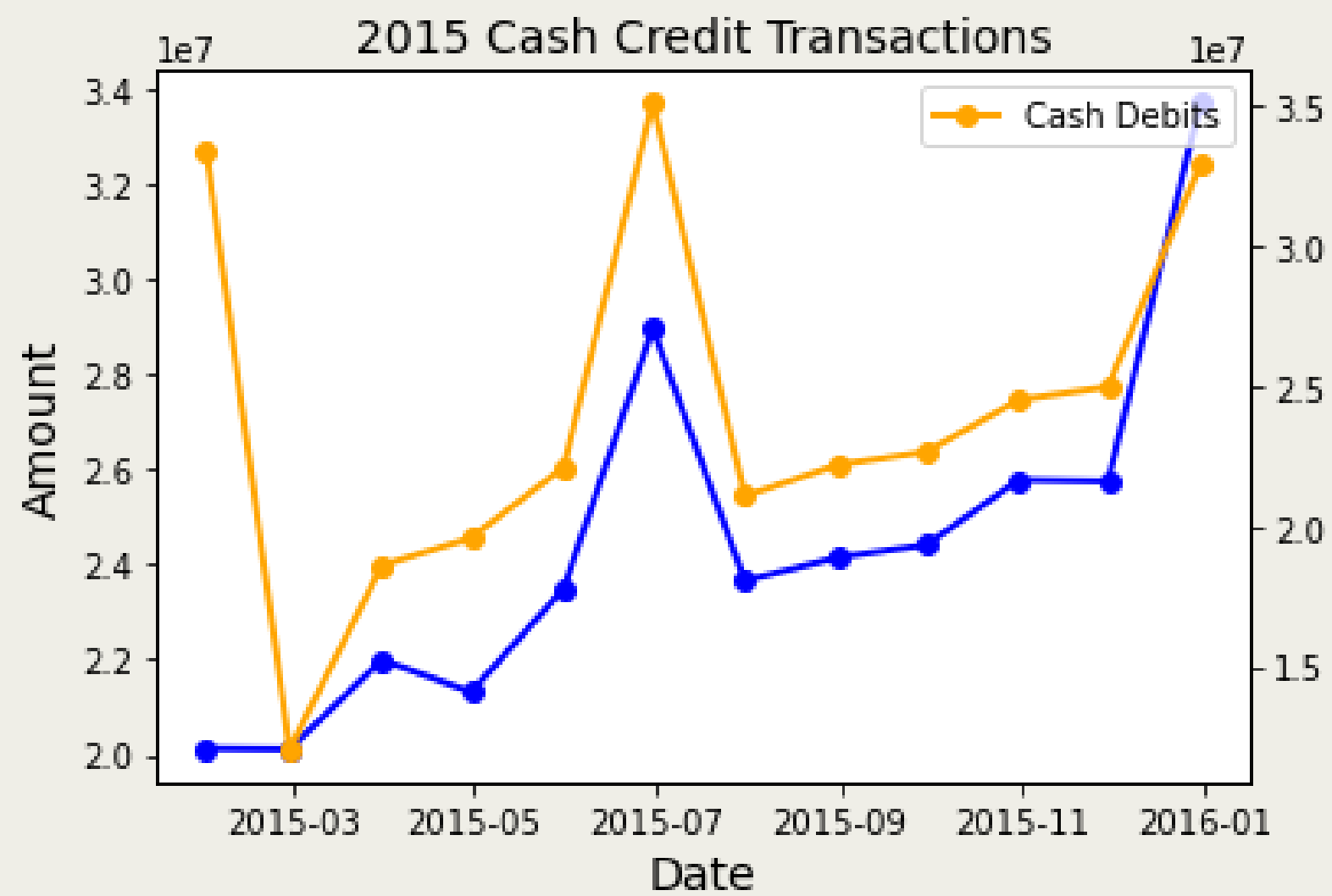
1. An upward linear trend - Cash deposits at the bank have gradually increased over the years.
2. Some seasonality is also observed - there is a downward spike in cash deposits at the beginning of the year.

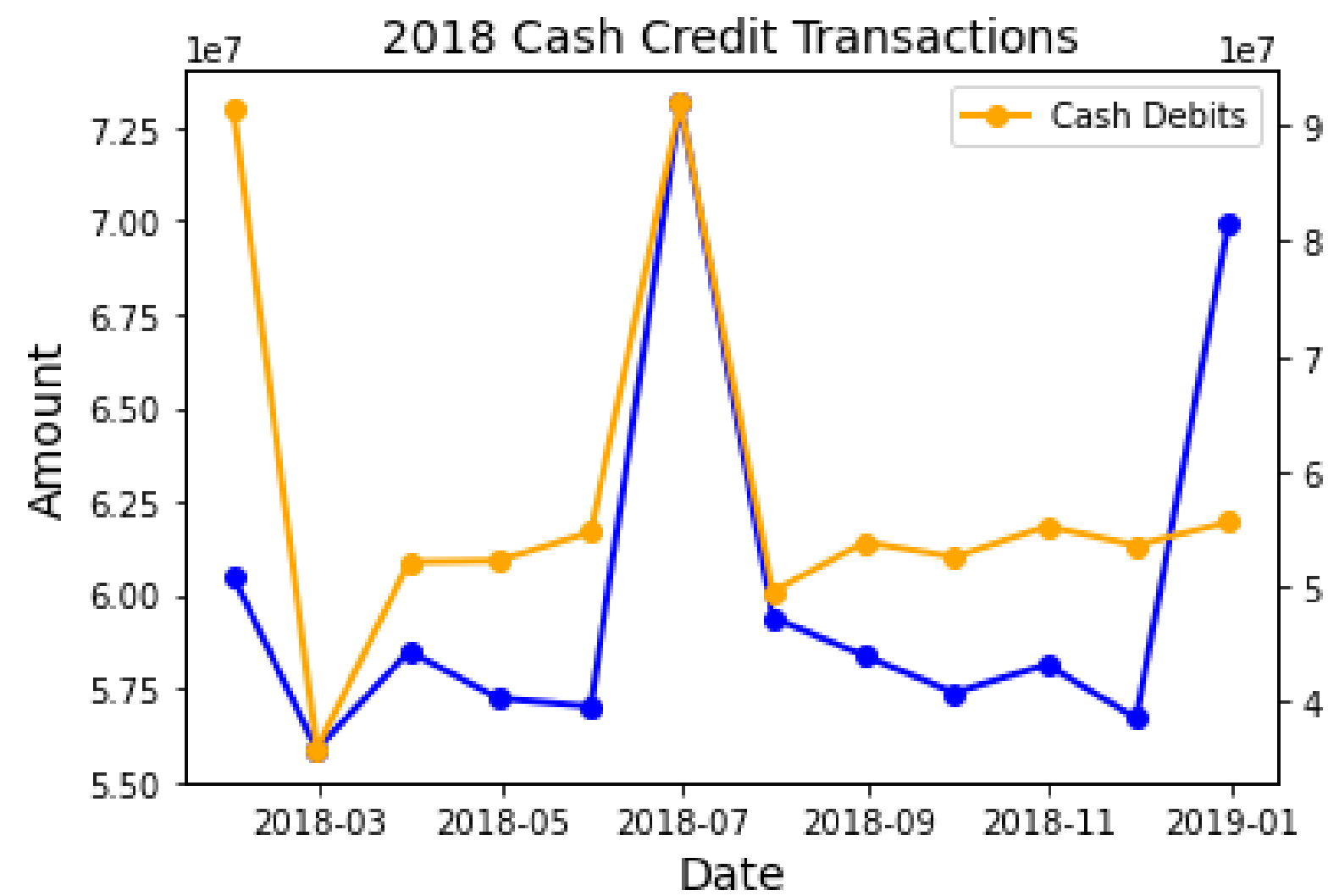
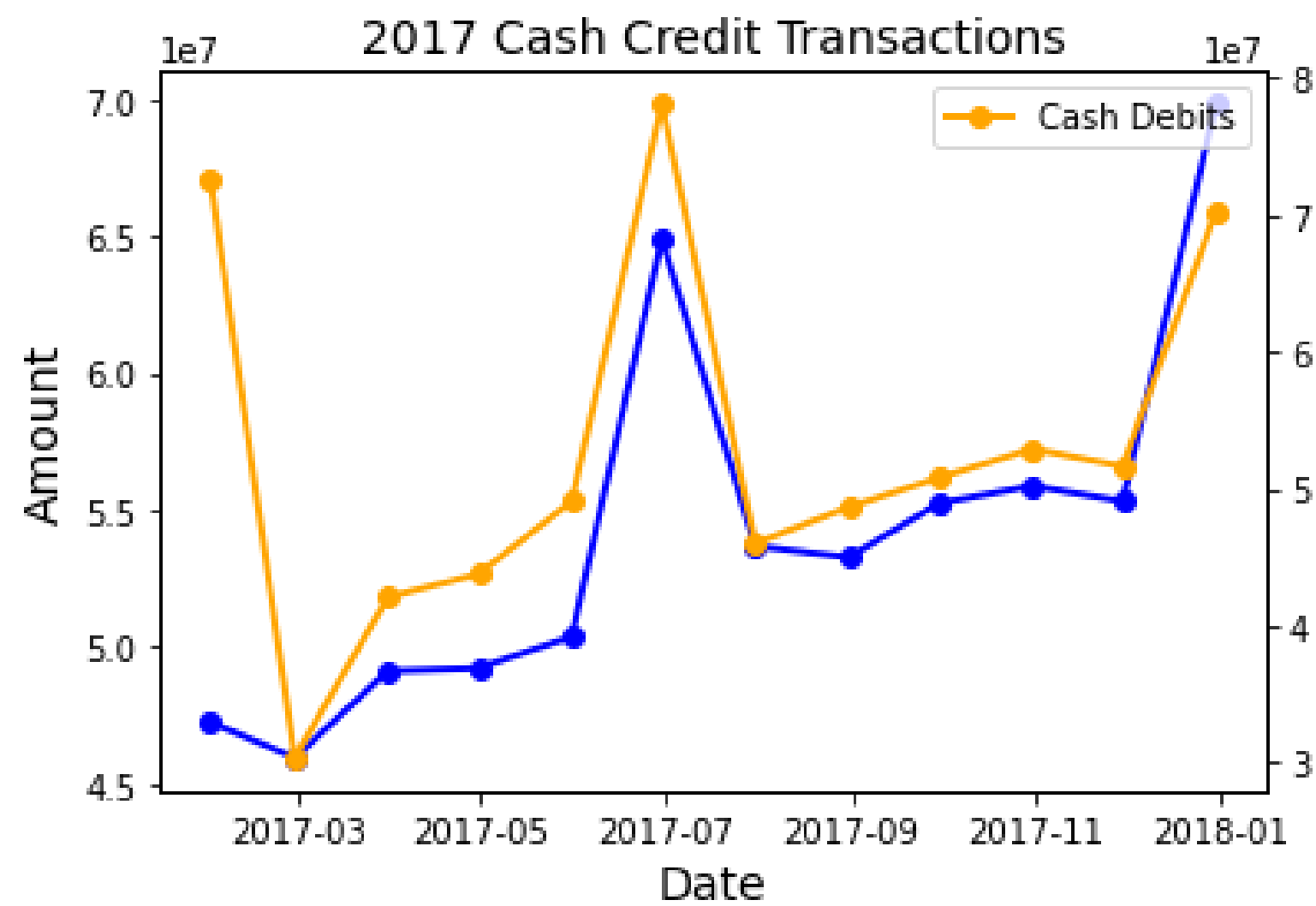


Observations:

1. An upward trend - Cash withdrawals at the bank have gradually increased over the year with periodic spikes at the beginning and mid-year.
2. A huge downward spike is observed towards the end of 2018 and the beginning of 2019. There was an increase in consumer prices and inflation in January 2019 which discouraged spending among the citizens hence cash withdrawals were made







Observations

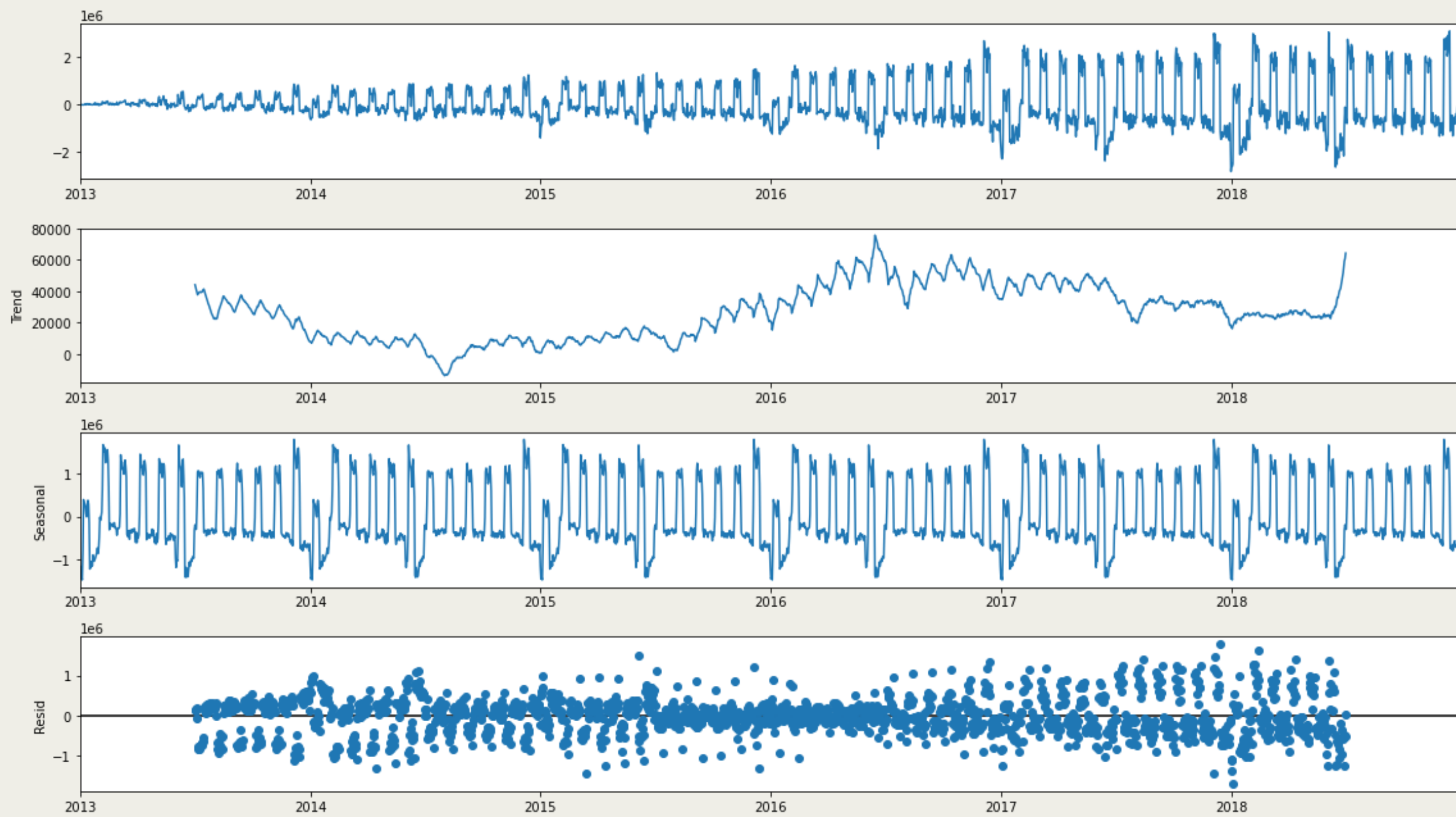
- More trends are observed from the yearly plots:
- There is an upward spike in cash withdrawals at the end of the year and the opposite for cash deposits(except 2013 which is the bank branch inception year.Very few transactions were also observed during this year.
- There is an increse in both credits and debits in the month of June - which marks the beginning of Summer in the region which attracts alot of tourists

PRE-PROCESSING

CLOSING BALANCE:
ADF STATISTIC: -5.87776291678782
P-VALUE: 3.130122828540417E-07
CRITICAL VALUES:
1%: -3.4333740509447717
5%: -2.8628759187993382
10%: -2.567481176655697

Checking for stationarity

The test statistic is less than the critical value at a significance level of 5% and the p-value is also less than the significance level of 0.05 hence we reject the null hypothesis and consider the series as stationary



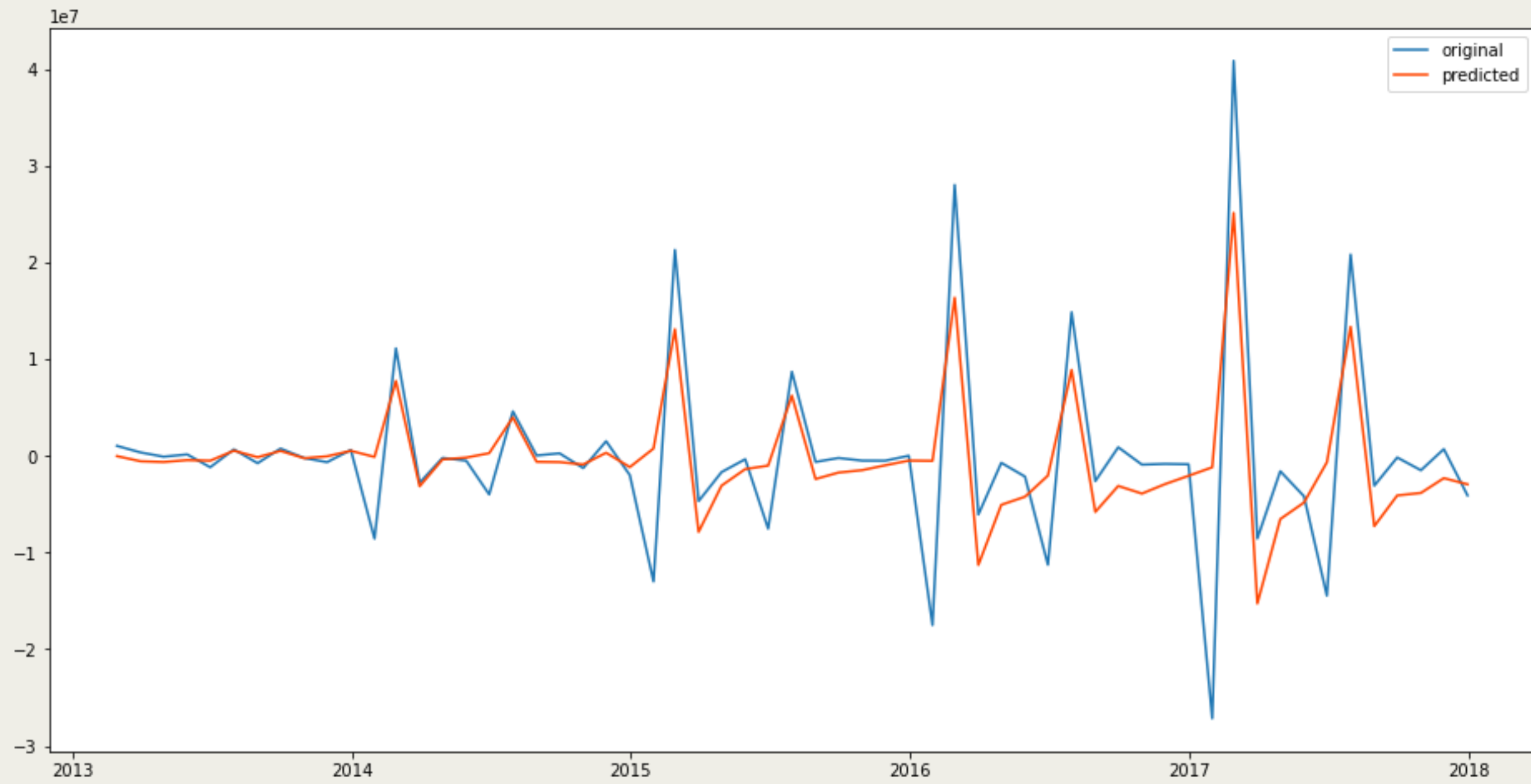
Observations:

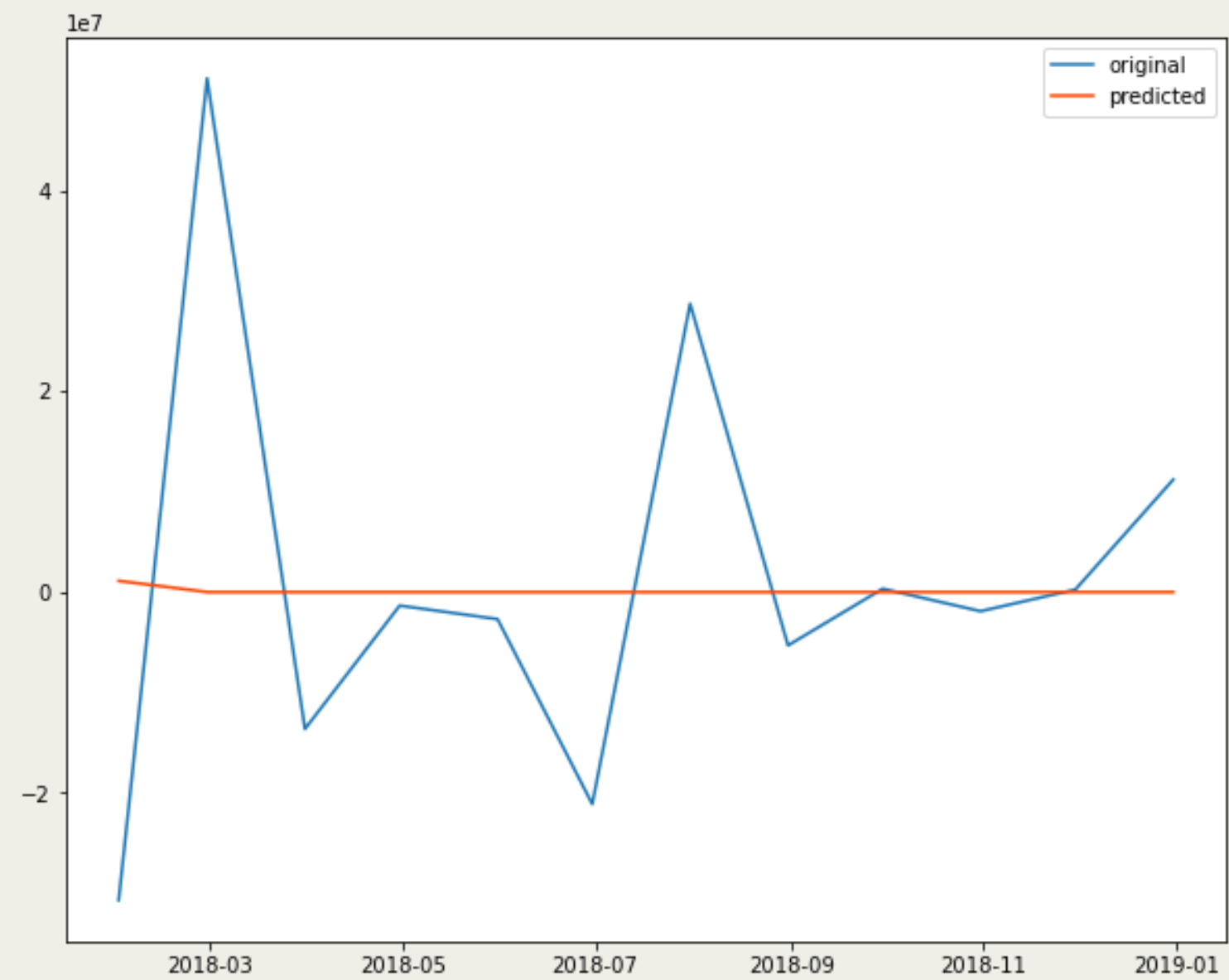
1. This is the original observed data that we are decomposing.
2. The trend component shows an upward direction. The trend indicates that there is a positive (increasing or upward) long-term movement.
3. The seasonal component shows the seasonality effect and the repeating pattern of highs and lows within the year.
4. Finally, the residual (noise) component shows the random variation in the data after applying the model. In this case, a multiplicative model was used.

MODELLING

We started by finding the autocorrelation. In order to find out which lags had a strong correlation to the forecasted value we used the `plot_acf()` function from the `statsmodels` module. This creates a plot that shows how much each previous lag influences the future lag.

MODEL 1: ARIMA

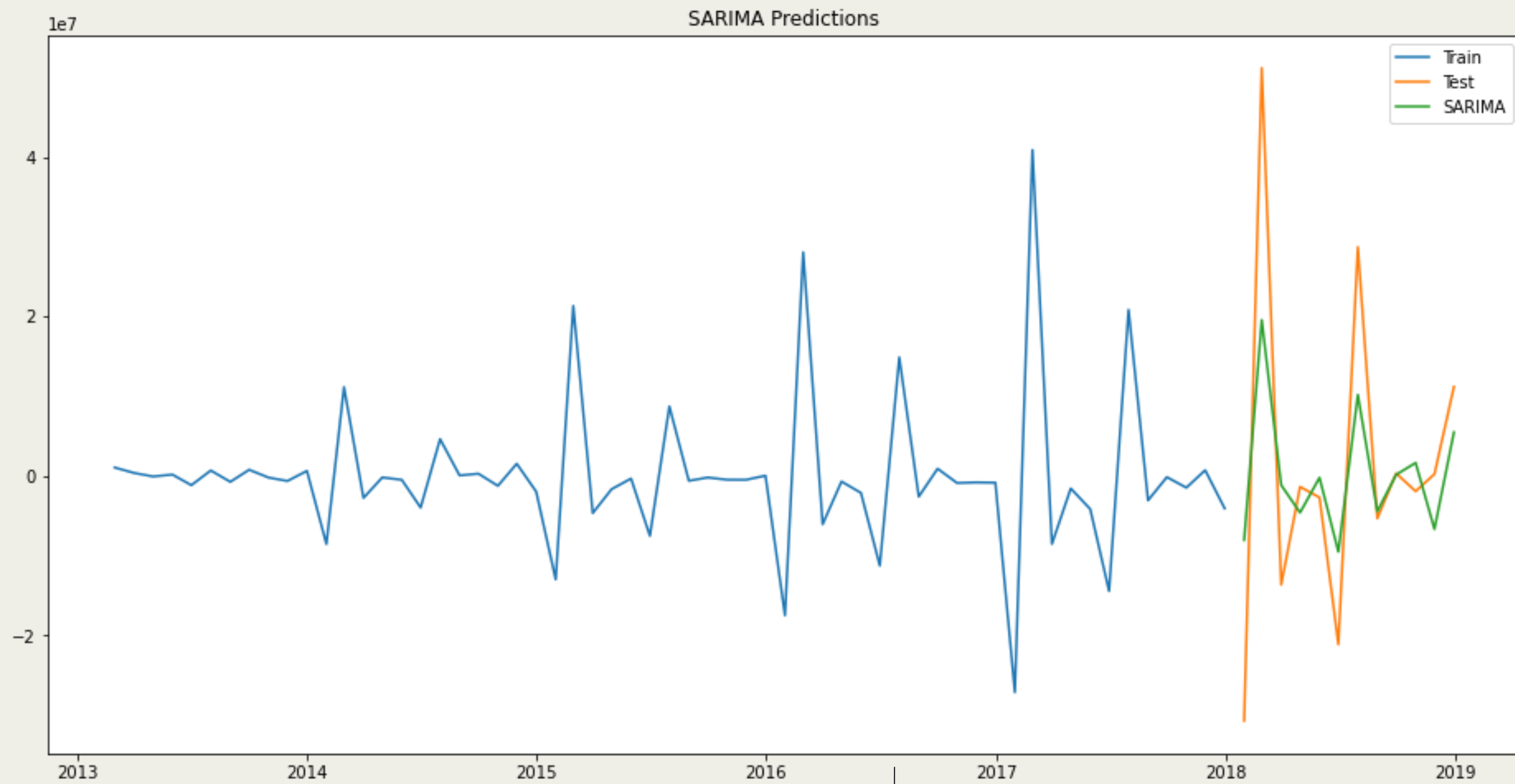




Observations:

1. The model overfitted on the train set and performed poorly on the test set

MODEL 2: SARIMA



EVALUATION

TEST RMSE: 13720318.067
TEST MAPE: 334.335
TEST MPE: 294.740

Model Performance Observations

1. ARIMA model - the model was overfitted on the train and performed poorly on the test data.
 -
2. SARIMA model - the model performed fairly well on both the train and test data. It captured the monthly seasonality in the cash balances well but did not capture the extreme values.
 -
3. Further attempts to reduce the variance and seasonality present in the data failed. The SARIMA model performed worse on the data after the second differencing.

CONCLUSION

1

There is seasonality in cash transactions - we observed spikes in cash withdrawals at the end of the year and in the mid-year around June. A high volume of cash credits is also observed in the mid and end-of-year periods(June and December)

2

There are more instances where the cash withdrawals outweighed the cash balances - indicating that the bank frequently requested for extra cash to cover heavy withdrawals

3

There was no observed correlation between a particular day of the week and the net cash balance.

4

There is predictability in the cash balances - our final model was able to capture correctly the monthly seasonality in the cash balances. The model, however, did not capture instances where there were extreme values.

RECOMMENDATION

1	2	3	4
<p>The bank should retain high cash balances during months where high volumes of withdrawals were observed (December and June) to be able to sufficiently cater to their customers.</p> <p>The bank should also maintain a high level of security in instances where high cash inflows are expected during this time.</p>	<p>The bank should employ our model to forecast expected future cash balances - our final model (SARIMA) was able to accurately capture the seasonality in cash balances.</p> <p>The bank could use this to prepare adequately for periods where cash balances are expected to be high by repatriating the bigger portion of their cash to the central vault as they expect high inflows of cash.</p>	<p>One of our findings was the bank was frequently caught off guard by high volumes of withdrawals and frequently had little cash to carry forward to the next day</p> <p>- by employing this forecast model, the bank should be able to maintain adequate balances within its premises.</p>	<p>The bank should use the seasonality captured by our model to come up with a repatriation schedule that can collect from branches experiencing high cash balances and redistribute to those experiencing the opposite.</p> <p>This will greatly cut down on operational costs that come with employing professional courier services to constantly deliver cash to and from the central vault.</p>

THANK YOU FOR
YOUR TIME!