(A903) Assignment: AI-Driven Predictive Cash Flow Analysis Using MT940 Messages

Problem Statement Description

- Analyze the structure and entries of MT940 messages.
- Extract relevant transactional data for cash flow analysis.
- Utilize AI to forecast upcoming bills and expected income, thus aiding in managing cash flow more effectively.

My proceedings

Preprocessing

- Assigning proper datatype to each field in the data.
- Grouping the data as per the owners (as in we need to check the cash flow for each company/owner).
- Removing the unwanted columns.
- Sorting the data according to the dates to understand the trends in the data according to the time.

About Data

- The whole data consists of data of various companies and usually the opening balance of the current day and the closing balance of the previous day must be same, but it is not seen in this case.

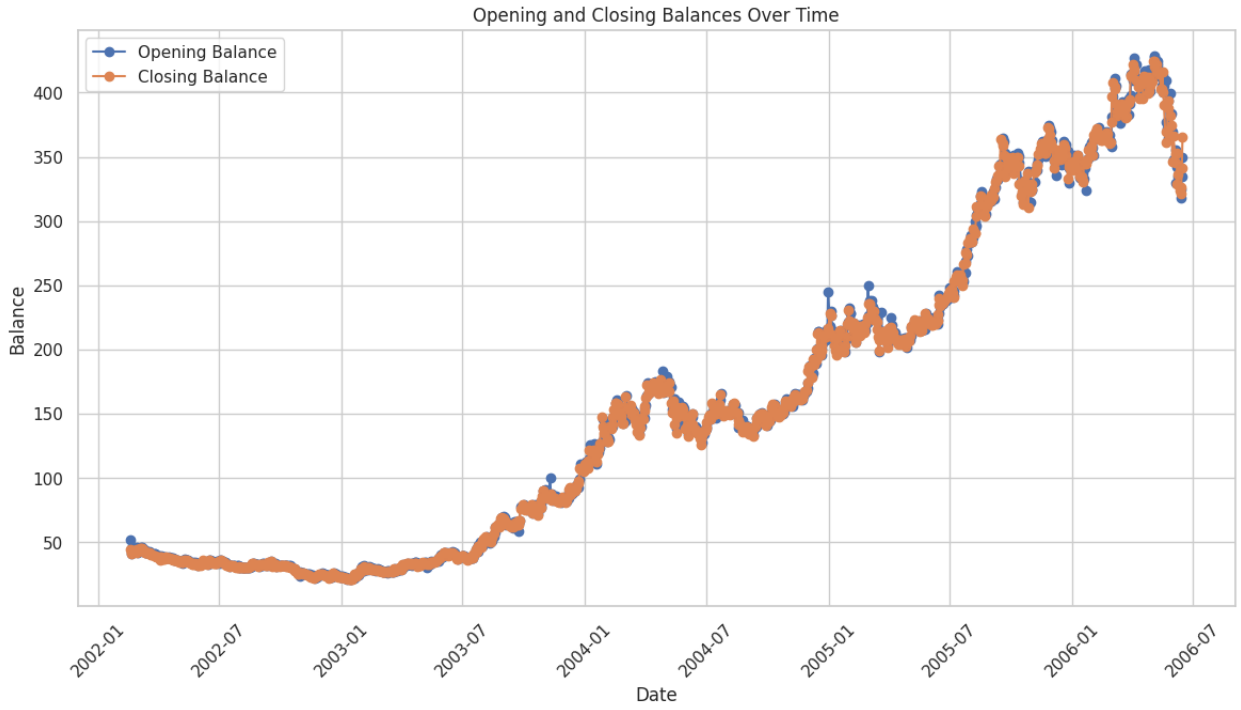
    a. First, I thought we should alter the data and adjust those balances to make it consistent but after analyzing the trends in the difference between opening balance of the current day and the closing balance of the previous day I noticed that those differences could impact the output so dropping the plan of

adjustments, conclusion from this was that there are transactions that are not recorded.

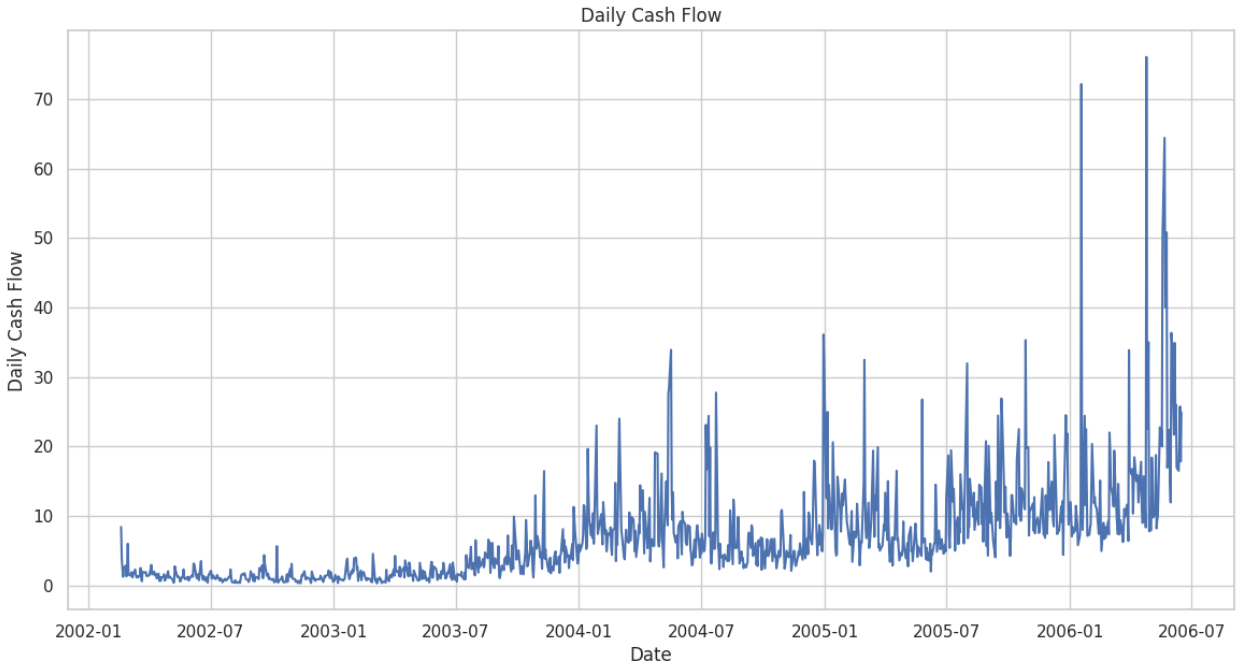
b. I have also not filled in the missing dates.

- I calculated the daily cash flow by subtracting debit from the credit.

Variation of Opening and closing balances with time.



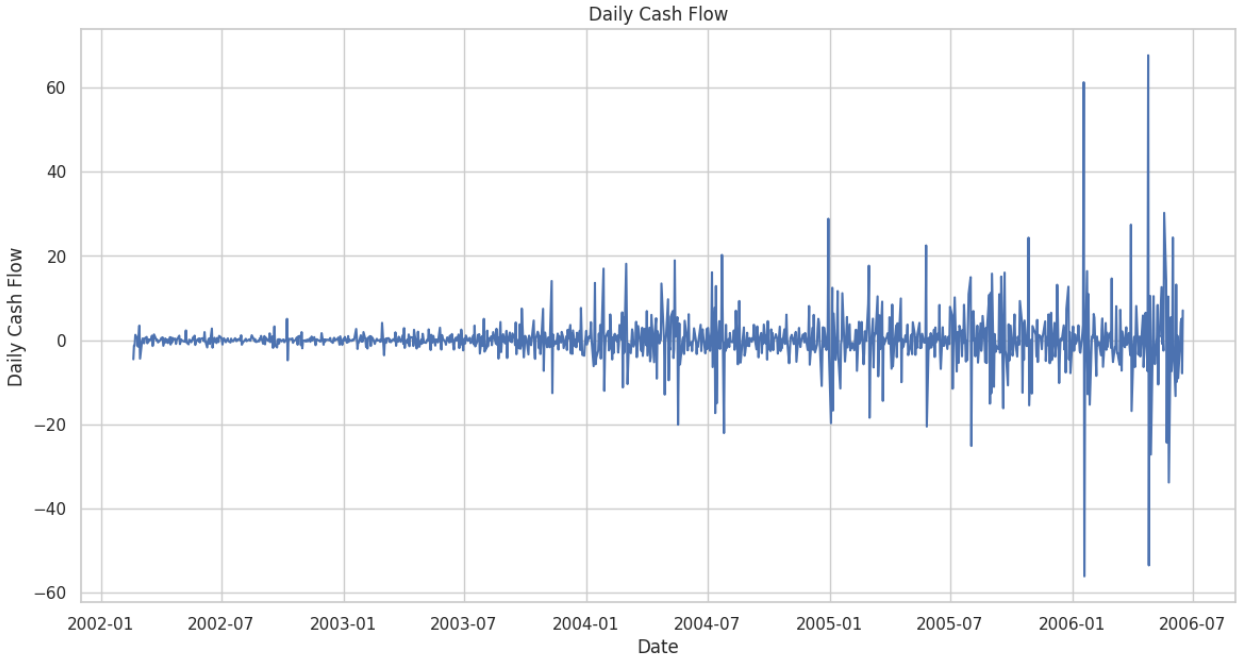
Variation of Daily Cash Flow with time.



Analysis

- Now as we can see the cash flow plot it is a bit difficult to guess whether the data is stationary or not so to check that there are various ways one of them Augmented Dickey-Fuller (ADF) test, the result of the test was the data is not stationary and for time series we require a stationary data, so we need to convert it to stationary by using differencing.
  But still, we can check the model's accuracy by fitting it with and without differencing.

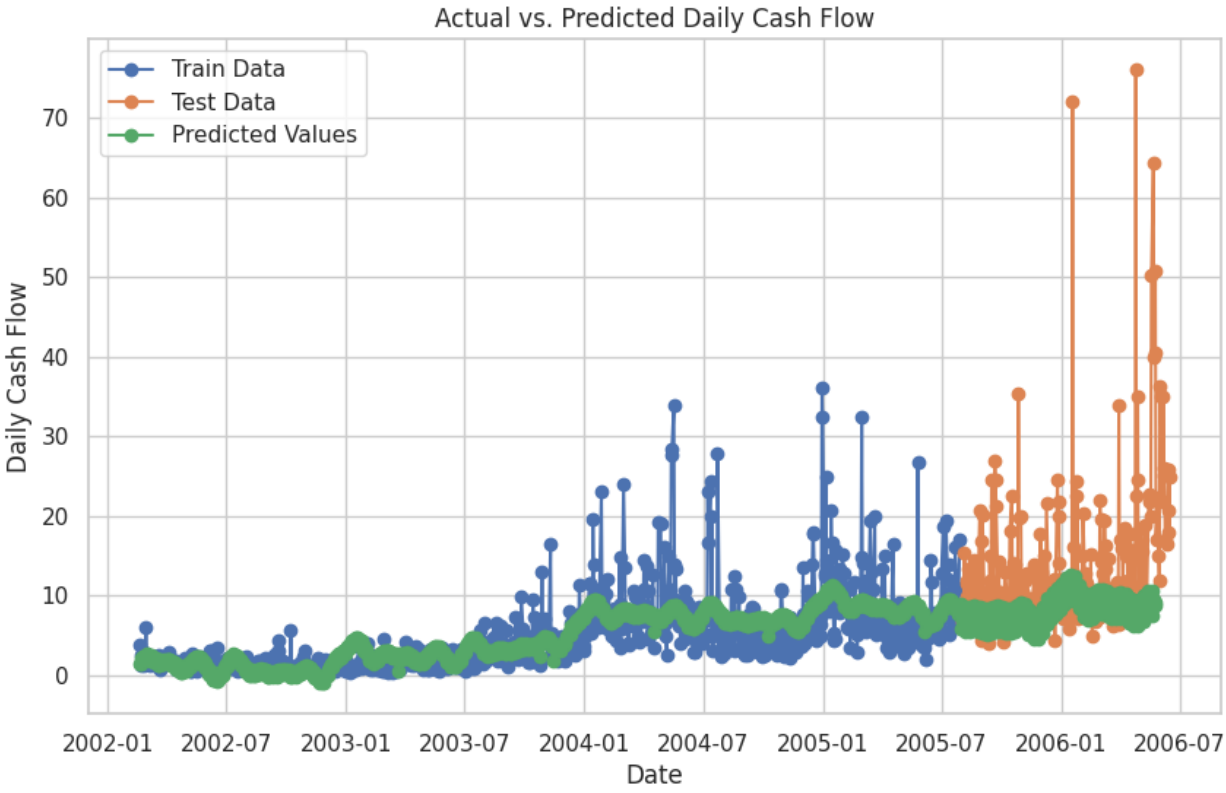
Daily Cash Flow with time after Differencing



- The next step was to identify TSA components in the data and by analyzing the graphs I noticed that seasonal component is present.

Model

- I have used a library from Facebook known as prophet and I trained the time series model using the data that was not differenced following is the graph of its performance.


Actual vs. Predicted Daily Cash Flow
Train Data
Test Data
Predicted Values
Daily Cash Flow
0
10
20
30
40
50
60
70
2002-01
2002-07
2003-01
2003-07
2004-01
2004-07
2005-01
2005-07
2006-01
2006-07
Date