Time Series Analysis Report: SBI ATM (On-Site and Off-Site) Forecasting

1. Introduction

This report documents the systematic time series analysis and modeling carried out for the "ATM(on-site)" and "ATM(off-site)" columns of the State Bank of India (SBI), using statistical modeling to produce forecasts and benchmark their accuracy.

2. Data Extraction and Selection

- Data for the State Bank of India was extracted from the master dataset by filtering rows where "Bank Name" includes "STATE BANK OF INDIA".
- Only the columns "ATM(on-site)" and "ATM(off-site)" were included for forecasting. All other columns—such as POS(Total), No. of Credit Card, No. of Debit Card, POS/QR/UPI metrics, and all transaction amount columns—were excluded from the core ATM analysis. The reason for exclusion is that these variables are either unrelated to the physical ATM network, represent different digital financial channels, or exhibit sporadic and missing data over the historical period.

3. Data Preparation and Cleaning

- The ATM columns were converted to numeric datatypes, and any parsing failures were handled as missing values.
- The dataset was checked for size and completeness. Exploratory plotting was performed to verify the overall quality and time continuity of the series for both ATM(on-site) and ATM(off-site).

4. Exploratory Data Analysis

- A joint plot of ATM(on-site) and ATM(off-site) over the entire available period revealed long-term trends and potential seasonality.
- Both series exhibited cycles and possible yearly effects.

5. Time Series Decomposition

- Seasonal decomposition was performed (additive model with a 12-period/month season) for each ATM type.

- The decomposition isolated the trend component (long-run change), the seasonal component (repeating yearly cycles), and the residual component (irregular short-run variation).

6. Stationarity Testing

- The Augmented Dickey-Fuller (ADF) test was applied to both ATM(on-site) and ATM(off-site).
- ATM(on-site) was found to be stationary (ADF p-value < 0.05), so it was suitable for direct SARIMA modeling without differencing.
- ATM(off-site) was found to be non-stationary (ADF p-value > 0.05). First-order differencing (subtracting each value from its preceding value) was performed, after which the series passed the stationarity test.

7. SARIMA Model Fitting and Forecasting

- For ATM(on-site), a SARIMA model was fitted with order (1,0,1) and seasonal order (1,0,1,12).
- For ATM(off-site), a SARIMA model was fitted to the differenced series with order (1,0,1) and seasonal order (1,0,1,12). The predicted differences were converted back to actual ATM counts by cumulatively summing the forecasted changes starting from the last observed value.
- Both models were used to forecast the next five months (Jan–May 2025).

8. Model Evaluation (SARIMA)

- The predicted values for Jan–May 2025 were compared to the actual numbers for both ATM(on-site) and ATM(off-site).
- Accuracy was quantified using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

9. Benchmarking with Alternative Models

9.1. Holt-Winters (Exponential Smoothing)

- Additive Holt-Winters models were independently fitted to both ATM(on-site) and ATM(off-site).
- 5-step forecasts and accuracy metrics were generated and compared to actuals.

9.2. VAR and VARIMA Models

- Vector AutoRegression (VAR) and Vector AutoRegressive Integrated Moving Average (VARIMA/VARMAX) models were fitted jointly on both ATM series where possible.
- 5-step forecasts were generated and evaluated against the same actuals.

Model	**ATM(on-site) MAE**	**ATM(on-site) RMSE**	**ATM(on-site) MAPE**	**ATM(off-site) MAE**	**ATM(off-site) RMSE**	**ATM(off-site) MAPE**
SARIMA	985.41	1005.63	3.61%	804.46	1148.15	2.20%
Holt-Winters	678.47	748.41	2.46%	1352.59	1599.70	3.85%
VAR	892.11	964.70	3.39%	1004.76	1302.73	2.53%
VARIMA	987.42	999.43	3.65%	1071.31	1202.41	2.86%

11. Conclusion

The time series analysis focused exclusively on the physical ATM deployment channels within SBI, in line with the study aim and due to data reliability in these columns. The analysis included visualization, decomposition, strict stationarity testing, SARIMA and multiple benchmark statistical models, multi-step forecasting, and comparison to actual observed values.

Across all models, the accuracy was assessed with standard industry scoring metrics over the most recent forecasting window. This pipeline is suited for ongoing monitoring, network planning, or performance analysis of SBI's ATM infrastructure.