Canadian Banks Stock Analysis Report
Authors: Mariia-Olena Zhupnyk & Mariia Shekhovtsova
Date: 17 Feb 2025

Canadian Banks Stock Data Analysis & Prediction Report

1. Introduction

1.1 Project Overview

The financial sector plays a crucial role in the Canadian economy, and the performance of major banks significantly impacts investors' decisions. This project aims to analyze and predict stock prices for five major Canadian banks: RBC, TD, CIBC, BMO, and Scotiabank. The analysis involves web scraping, data storage, exploratory data analysis (EDA), and predictive modeling using time series forecasting techniques.

1.2 Goals and Purpose

Before diving into the data, we defined the key objectives of our analysis by asking critical questions:

- **Trends & Patterns**: How have stock prices evolved over time? Are there noticeable seasonal or cyclical trends?
- **Volatility**: Which bank's stock is the most volatile, and which is the most stable?
- **Correlations**: Do the stock prices of these banks move together? How strong is their correlation?
- **Market Events Impact**: How do macroeconomic factors (e.g., interest rates, GDP, inflation) affect these banks' stock prices?
- **Comparative Performance**: Which bank performed the best over a given period (e.g., last year)?
- **Prediction**: Can we forecast future stock prices based on historical data?

1.3 Objectives

- Collect, clean, and store historical stock data remotely.
- Conduct Exploratory Data Analysis (EDA) to identify trends and patterns in stock movements.
- Implement time series forecasting techniques to predict future stock prices accurately.
- Compare different banks' performances to evaluate investment opportunities.

Canadian Banks Stock Analysis Report
Authors: Mariia-Olena Zhupnyk & Mariia Shekhovtsova
Date: 17 Feb 2025

2. Environment Setup

As part of the data engineering phase, establishing a robust infrastructure was essential. The project stores stock data in a remote PostgreSQL database. However, connecting to the server posed a challenge due to its dynamic IP address. After thorough research, **ZeroTier** was implemented as a networking solution, allowing seamless access to the database. This setup ensures stable connections, enabling efficient data retrieval, transformation, and analysis.

screenshoots

ДОПИСАТЬ

3. Data Collection & Storage

Data Sources:

- Stock price and financial data were web-scraped from Yahoo Finance using Python.
- The data includes stock price history, financial statements (income statement, balance sheet, cash flow), and dividends.

Database Setup:

- PostgreSQL was chosen for storing structured stock data.
- A remote server hosts the database, enabling collaboration and secure data access.
- Data is stored in tables categorized by stock price history, financial statements, and computed technical indicators.

---

2. Data Collection & Processing

2.1 Data Source

The dataset was obtained using **web scraping from Yahoo Finance** and includes stock prices from **January 2020 to December 2024**.

Web scraping is an efficient and automated method to gather data from the internet, ensuring we have up-to-date and comprehensive datasets.

2.2 Data Structure

Canadian Banks Stock Analysis Report
Authors: Mariia-Olena Zhupnyk & Mariia Shekhovtsova
Date: 17 Feb 2025

The dataset consists of the following columns:

- stock_ticker – Bank stock symbol (e.g., BMO.TO, TD.TO)
- date – Trading date
- close_price, high_price, low_price, open_price – Daily stock prices
- volume – Number of shares traded
  These are essential variables for understanding the price dynamics and trading activity.

Feature Engineering

- **Simple Moving Averages (SMA):** SMA_50, SMA_100, SMA_500 – These represent the average closing price over the past 50, 100, and 500 days.
- **Exponential Moving Averages (EMA):** EMA_50, EMA_100, EMA_500 – These are more responsive to recent price changes.
- **Daily Return (daily_return):** Measures the percentage change in the closing price compared to the previous day.
  These features help identify trends, smooth out fluctuations, and assess risk, improving the quality of the analysis and forecasting models.

2.3 Data Cleaning

- Handled missing values.
- Converted data types for analysis.
- Calculated new features such as **moving averages** and **daily returns**.

---

3. Exploratory Data Analysis (EDA)

EDA is a crucial step in the data analysis process, including when dealing with transactional data. EDA helps to gain insights, identify patterns, understand the structure of the data, and uncover potential issues.
EDA ensures that the data is understood fully, cleaned effectively, and prepared for deeper analysis or predictive modeling.

- **Descriptive Statistics & Data Summary:**
  Calculated key statistical measures (mean, median, standard deviation) to understand stock price distributions.
  This provides basic insights into the data's spread and central tendencies, crucial for detecting anomalies.
- **Trend Analysis & Moving Averages:**
  Computed Simple Moving Averages (SMA) and Exponential Moving Averages (EMA) to

identify long-term trends.
These techniques help smooth out price fluctuations and identify long-term market direction.

- **Volatility & Risk Analysis:**
  Measured daily returns for each stock to assess risk levels.
  Volatility is an important metric for understanding the stability of stock prices and the risks involved.
- **Correlation Analysis:**
  Calculated correlations between different bank stocks to determine how they move in relation to each other.
  Understanding stock correlations helps in assessing portfolio diversification and identifying potential risks and opportunities.
- **Market Event Impact:**
  Explored how macroeconomic factors influence stock prices.
  This provides a broader context for understanding the external factors that impact the banks' performance.
- **Visualization Techniques:**
  Used Matplotlib and Seaborn for static plots.

  Visualizations make it easier to identify patterns, relationships, and trends, providing clearer insights.
  -

3.1 Correlation Analysis

- A correlation matrix was computed to identify relationships between stock prices of different banks.
- Strong positive correlations (>0.85) were observed, indicating that banks' stock prices tend to move together.

3.2 Visualizations

- **Pairplots and scatter plots** were used to analyze stock price distributions and relationships.
- **Time series plots** showed long-term trends and volatility.

TODO busines insights

Identify key questions to answer like: Which bank has been the most volatile? Which bank has had the highest average return?

Canadian Banks Stock Analysis Report
Authors: Mariia-Olena Zhupnyk & Mariia Shekhovtsova
Date: 17 Feb 2025

4. Predictive Modeling: Time Series Forecasting

4.1 Model Selection

We explored different forecasting methods:

- **ARIMA (AutoRegressive Integrated Moving Average)**
- **LSTM (Long Short-Term Memory) Neural Networks**
- **Monte Carlo Simulation** for risk assessment

4.2 Results

- **ARIMA provided stable forecasts** for short-term trends.
- **LSTM captured non-linear patterns** but required more tuning.
- **Monte Carlo simulations estimated stock price distribution** for different market scenarios.

TODO

Compare Models:

- Evaluate the performance of ARIMA vs. Monte Carlo simulation. Which model gives you more accurate predictions?

---

5. Investment Insights

5.1 Performance Comparison

- **TD and RBC showed the highest average returns** over the analyzed period.
- **BMO and Scotiabank had more volatility**, increasing risk.
- **Moving Averages and Daily Returns** suggested stable trends for **TD and RBC**.

5.2 Recommendation

Based on **historical performance, trend stability, and risk factors**, **TD and RBC appear to be the better investment options** for long-term gains.

Canadian Banks Stock Analysis Report
Authors: Mariia-Olena Zhupnyk & Mariia Shekhovtsova
Date: 17 Feb 2025

---

6. Conclusion & Future Work

6.1 Summary

This project provided insights into Canadian banks' stock performance using data analysis and time series forecasting. Key findings include:

- Strong correlations among major banks.
- Stable upward trends in **TD and RBC** stocks.
- Time series forecasting indicating continued growth potential.

6.2 Future Enhancements

- **Enhance forecasting models** using hybrid approaches.
- **Incorporate economic indicators** like interest rates and GDP.
- **Automate real-time stock data analysis** for dynamic insights.

---

7. References

- Yahoo Finance API
- Pandas, NumPy, Matplotlib, Seaborn
- Statsmodels, Scikit-learn, TensorFlow

GitHub Repository: https://github.com/helenzhupnyk/canadian_banks_stock_data