Chantal Villacres
Xiashu Chu
Weisui Xu
Jun Lee
Liam Hasandjekaj

Baruch Capital's OTC Pricing Pipeline

Business Problem

Baruch Capital is an investment firm looking to gain a competitive advantage in its industry by gaining increased insight and analytical capabilities into the OTC market. The Over-the-Counter (OTC) market is a decentralized marketplace where securities are traded directly between two parties without the traditional supervision of an exchange such as the NYSE or Nasdaq. The OTC market works through a network of dealers and brokers who facilitate transactions "over-the-counter." Here buyers and sellers can negotiate directly with one another to execute trades, typically through phone, email, and online messaging through applications such as Bloomberg. This allows for high flexibility and customization in trading terms with respect to price, quantity, and settlement dates, making it a highly attractive space for Baruch Capital to trade the types of securities that might not meet the listing requirements of formal exchanges. Baruch Capital is looking to gain an edge against its peers by building a data pipeline which gathers external OTC pricing data and presents relevant technical indicators to its analysts and traders through a user-friendly dashboard. Given the decentralized nature of OTC transactions, well-informed, data-driven insights are difficult to obtain. With this new data pipeline project, Baruch Capital aims to establish a data pipeline that ingests external OTC pricing data, stores it in an efficiently built data warehouse, and serves the data to relevant stakeholders through a business intelligence dashboard that can support key technical indicators and visualizations. Company traders will use this platform to make better investment decisions in the highly competitive and lucrative OTC market.

Business Requirements

For this project, several core requirements have been identified in conversation with managing directors, traders, and data analysts at Baruch Capital. Firstly, the OTC pricing dashboard must provide a comprehensive view of a range of technical indicators identified by the firm's analysts. These technical indicators should display a range of time frames, including 3 months, 6 months, 1 year, 3 years, and 5 years for each security in the OTC market data pipeline. In addition these technical analyses should be automatically updated based on the current date, ensuring relevance and accuracy. The dashboard should also offer drill down views by year, quarter, and month, in order to facilitate a deeper analysis of OTC trends. The project also requires segmentation of security data by country, industry, security type, and additional relevant

fields to be determined by stakeholders. By meeting these requirements, Baruch Capital aims to empower its analysts and traders with actionable insights derived from an OTC market data pipeline.

Business Impact

As with any new development initiative, the OTC pricing pipeline project has potential risks, costs, and benefits that must be considered. The major risks with this project are based on the potential for bad or missing data. If traders are making decisions on what to buy and sell with either inaccurate or incomplete data, then there is significant potential for large financial losses as a result of this project. However, the potential for systems failure is always present in a project as data-intensive as this one. Of course, there are also costs associated with the actual implementation of this pipeline such as labor and consulting costs for the data engineers to build the pipeline, maintenance costs for upholding the system throughout its lifecycle, and of course storage costs for the massive data warehouse.

However, the potential benefits of this project are massive for Baruch Capital. By providing analysts with access to comprehensive and real-time pricing data, the OTC pipeline project will allow traders to enhance their decision-making capabilities within a highly competitive market space where data-driven insights are invaluable. Baruch Capital will be better able to identify lucrative trading opportunities, optimize trading strategies, analyze market trends, and mitigate risks with the centralized OTC pricing data in hand provided by this system. With improved insights into market trends, patterns, and volatility, Baruch Capital can react swiftly to market changes, gaining an edge in a fiercely competitive industry. The technical analysis aspect of this project will also be a source of financial gain for Baruch Capital as the system will serve easy to read and accurate technical indicators to key decision makers. The ability to break down historical data, execute technical analysis, and determine the best actionable insights will allow traders to execute trades that are more favorable for the company.

A "moonshot" analysis of this system, if things go exactly as planned, would include Baruch Capital benefiting immensely from the efficiencies gained through these increased data insights. The firm would increase its gains, minimize its losses, reduce its operational costs, gain reputational credit, and increase its market share greatly. Ultimately, the OTC pricing pipeline project has the potential to revolutionize Baruch Capital's trading operations, driving explosive growth and profitability.

Business Persona

Traders: The primary beneficiary from the implementation of this system are the company traders who will utilize the OTC pricing pipeline's insights to make better investment decisions with the aim of maximizing profitability.

Analysts: The technical indicators and data visualizations served by this pipeline will allow company analysts to perform technical and predictive analyses as well as identify new market opportunities to support the traders.

Data Engineers: The company's data engineers will enhance and maintain the OTC pricing pipeline system over its lifecycle after ACE Consultant creates the system, ensuring that the data is accurate, useful, and satisfies the requirements as identified by analysts and traders.

IT: The information technology team will provide technical support for connectivity, set-up, and configuration of the system for us by traders and analysts.

Data Source

The external data source for the OTC pricing pipeline project will be otcmarkets.com. The pipeline will extract data from this website, load that raw data into a cloud storage system and database, transform the data based on the requirements and data constraints of the company, and then serve it to traders and analysts through a business intelligence software platform. The first dataset we used for this project was 16gb. The dataset provided all the necessary fields however it did not have a readily available unique identifier for each record. Additionally many of the data types and formats needed to be altered significantly so as to fit with the requirements. Additionally, the dataset being very large proved to be problematic in loading all at once and required batch loading. We also had to thoroughly clean the dataset, removing null values, null columns, and duplicate records.

Methods, Data Tools, & Interfaces

After getting familiar with the external data source and the relevant columns of the dataset being used, ACE Consultants began building an information architecture. The team utilized Dbschema, a database schema design tool, in order to develop a dimensional model and set up the necessary tables, attributes, and relationships for the OTC pricing dataset. A snowflake schema was used for this project. It is defined by the date, security, status, tier, venue and location dimensions with a facts dimensional table in the center connecting them all. In order for the team to work together and handle push/pull version control, communication, and task assignments, ACE Consultants utilized GitHub repositories for team collaboration. The team primarily coded in Visual Studio Code in conjunction with Google Collab for testing the relevant data pipeline scripts. The data was stored in Azure Blob, Microsoft's scalable cloud storage platform due to its ease of use and scalability. For the pipeline methodology we utilized an extract, load, transform method (ELT). Given the large size of the dataset, the team chose ELT

for its superior scalability, faster data loading, and flexibility with respect to transforming the data directly within the data warehouse. Dbt in conjunction with DataGrip was utilized to transform the dimensional tables and data within the data warehouse. PostgreSQL was used as the database management system. Finally, Tableau was used as the end user interface that analysts and traders would use to easily access the data visualizations and business intelligence dashboards.

Database Management: DbSchema, DataGrip, PostgreSQL
Programming: Python, SQL
Development & Data Engineering: Visual Studio Code, dbt
Cloud Storage: Azure Blob
Data Visualization and Analysis: Tableau