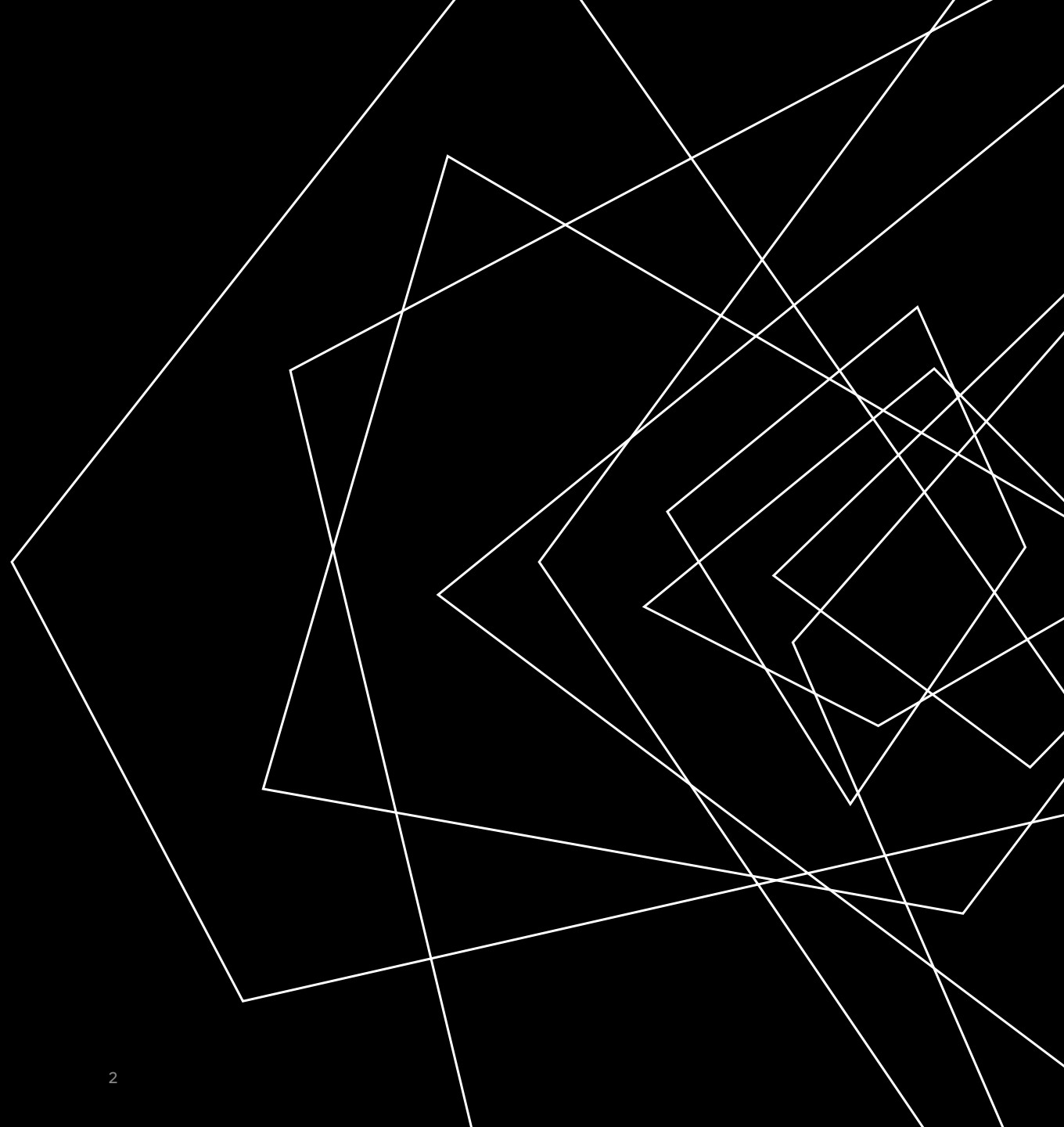
DATA ANALYSIS AND DESIGN

Steffen Anderson

7/15/2023

DATASET LIST AND ANALYSIS REPORT



INCITE

INCITE

INTRODUCTION

OVERVIEW

'Incite' will harness the power of data from Alpha Vantage, a reputable API service providing an array of real-time and historical stock data, including:

- Daily stock prices
- Qualitative and Quantitative indicators (Economic and Technical)
- Sector performances, and more.

This report will assess Alpha Vantage's datasets for the 'Incite' project, which aims to provide algorithmic trading models for retail investors.

https://www.alphavantage.co/

DATASET DETAILS

INCITE

By integrating these datasets, 'Incite' can offer retail investors comprehensive, data-driven tools for algorithmic trading that leverages the same sophistication as Wall Street strategies. This data allows 'Incite' to offer a rich, multifaceted approach to investment decision-making, addressing the key challenges faced by retail investors.

Core Time Series Stock Data APIs	Alpha Intelligence™	Fundamental Data	Indicators
- Offer granular time-series data of global equities at varying time intervals (intraday, daily, weekly, and monthly). - Key metrics provided include open, high, low, close prices and volume. - Useful for understanding price movements and volume trends over time.	- Comprises of News & Sentiment data, highlighting key news and sentiment scores related to various stocks. - Provides lists of Top Gainers, Losers, and Most Actively Traded Tickers in the US Market. - Ideal for capturing market sentiment, monitoring market movers, and identifying high-activity stocks.	- Provides critical company-specific data like earnings, income statements, balance sheets, and cash flow. - Enables deep analysis of company financials and identification of potentially profitable investments based on fundamental metrics.	Economic: - Offers macro-level data such as FX rates, crypto FX rates, sector performances, etc. - Helps understand broader market trends and economic conditions which could influence stock prices. Technical: - Includes a broad array of technical indicators such as - Simple Moving Average (SMA) - Exponential Moving Average (EMA), Weighted Moving Average (WMA), - Double Exponential Moving Average (DEMA), etc. - Supports the development of sophisticated, data-driven trading strategies based on established technical analysis principles.

INCITE

DATASET ASSESSMENT

DATA GAPS

- Availability and completeness of data for all companies in all time periods could be a challenge. Some financial indicators may not be available for all companies. Also, sentiment data could be biased towards larger, more popular stocks.

- Alpha Vantage provides a comprehensive suite of data for the stock market. However, while Alpha Intelligence™ includes some sentiment analysis data, it might not be as comprehensive or as granular as some projects might require. This could potentially be supplemented with other sentiment analysis sources for a more rounded view.

DATA QUALITY ASSESSMENT

- Historical stock price and financial indicators data are typically reliable and clean. However, news and social media sentiment data could be noisy and may require significant preprocessing and cleaning.

- Alpha Vantage is recognized for providing high-quality, accurate data. However, occasional missing data points or API limits could require additional data handling or management strategies.

RISK AND CHALLENGES

- While Alpha Vantage's data quality is high, the service's API has rate limits, which might restrict the amount of data that can be retrieved in a specific time period. Proper planning and efficient data fetching, and storage processes are essential to ensure smooth and continuous data flow for 'Incite'.

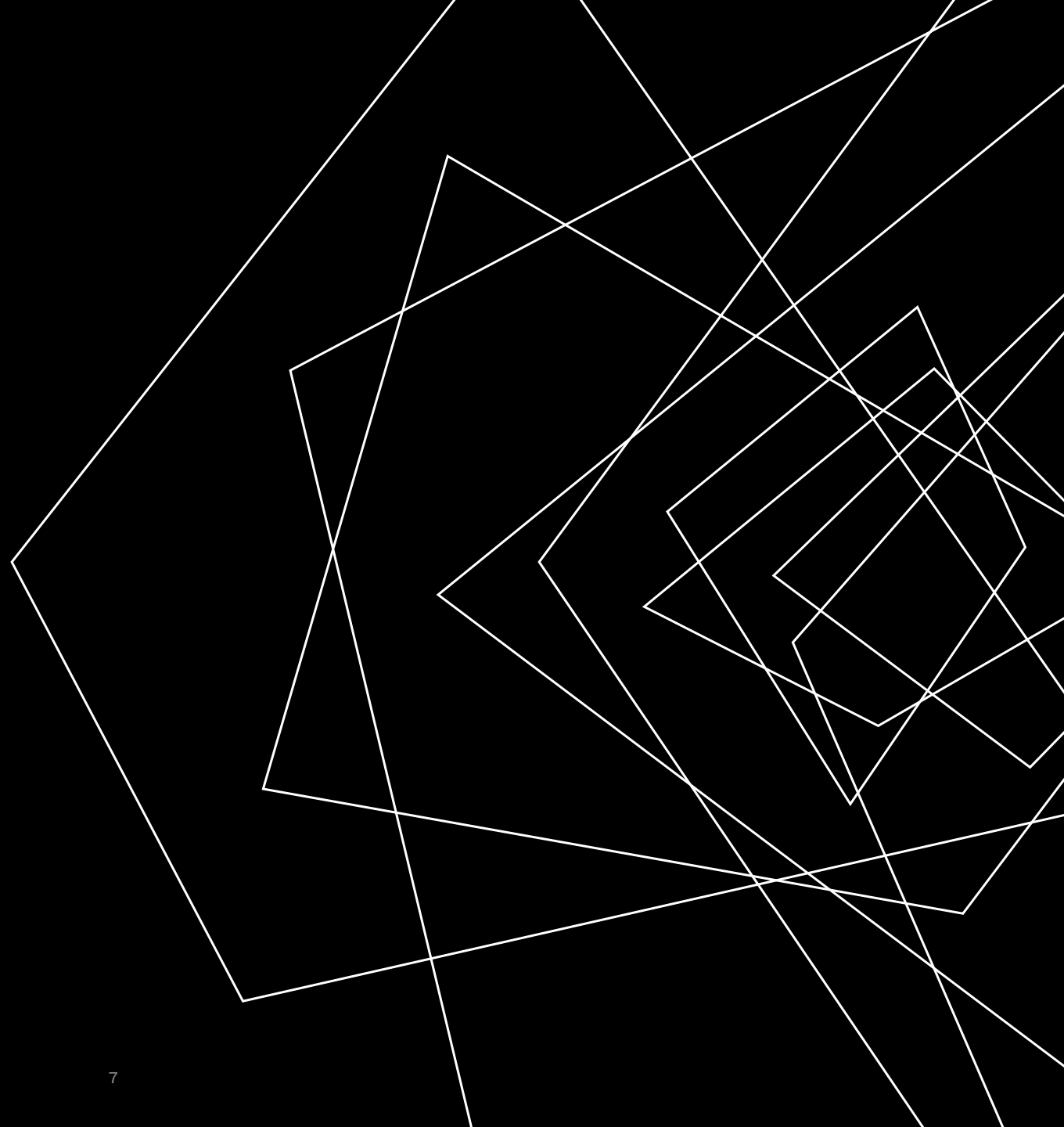
INCITE

CONCLUSION

KEY TAKEAWAYS

Despite minor challenges related to rate limits and potential data gaps, Alpha Vantage offers a solid data foundation for 'Incite'. Its rich suite of data, ranging from fundamental company data to technical indicators and sentiment analysis, will fuel the algorithmic trading models we're building. With continuous data validation and careful project planning, 'Incite' will democratize access to sophisticated trading strategies for retail investors. Compliance with Alpha Vantage's terms of use and data privacy regulations will ensure the project's long-term sustainability.

INSIGHTS AND TREND ANALYSIS REPORT



INCITE

INCITE

INTRODUCTION

OVERVIEW

The 'Incite' project aims to bring algorithmic trading to retail investors, leveraging robust datasets provided by Alpha Vantage. This report presents the key insights and trends that were identified during the initial stages of data exploration.

TRENDS IDENTIFICATION

In this rapidly changing financial landscape, it is essential to identify and understand current trends that can impact our algorithmic trading project. We've highlighted three key trends—increased market volatility, the rise of AI and machine learning in finance, and the surge in retail investor activity—that shape today's financial markets. These trends offer both challenges and opportunities that our model can address, laying a strong foundation for the 'Incite' platform.

TREND 1

Market Volatility: Recent years have seen increased market volatility, presenting both challenges and opportunities for algorithmic trading.

TREND 2

AI and Machine Learning in Finance: The finance sector, particularly investing, has seen a significant uptick in the use of AI and machine learning for predictive analysis and automation.

TREND 3

Retail Investor Activity: There has been a surge in retail investor activity, particularly in electronic and app-based trading platforms.

INCITE

INSIGHTS DISCOVERY

Delving deep into our initial data exploration, we've gathered valuable insights to guide our strategy for the 'Incite' platform. Our discoveries underline the importance of data-driven trading, the potential of machine learning, and the growing demand for sophisticated trading tools among retail investors. These insights not only validate our approach but also underscore the immense potential of our mission to revolutionize retail investing.

INSIGHT 1

Data-Driven Trading: Our initial exploration supports the idea that algorithmic, data-driven trading strategies can help retail investors make informed decisions and potentially yield better returns.

INSIGHT 2

Machine Learning Potential: Machine learning can harness market volatility to create trading strategies that adapt to market changes.

INSIGHT 3

Retail Engagement: With an increase in retail investor activity, there's an unmet demand for sophisticated, yet user-friendly trading tools.

INCITE

INCITE

DATASET ASSESSMENT

DATA SUFFICIENCY

- Alpha Vantage provides a wide array of data ranging from core time series stock data, technical indicators, fundamental data to economic indicators. This data is sufficient to develop and support algorithmic trading strategies and aligns well with the project goals.

CORRELATION ANALYSIS

- Our preliminary correlation analysis between different variables in the data supports the initial project charter. It indicates potential predictive relationships between various indicators and stock price movements.

ALTERNATIVE CONSIDERATIONS

- While Alpha Vantage's data is comprehensive for the current scope, expanding the project to include other data sources or types (e.g., more granular sentiment analysis) could be beneficial for more complex models in the future.

INCITE

CONCLUSION

KEY TAKEAWAYS

The insights and trends observed align well with the 'Incite' project objectives. The data from Alpha Vantage is comprehensive and supports the development of algorithmic trading models. As the project evolves, additional data sources or types could be considered to enhance model performance further. Nonetheless, 'Incite' is well-positioned to deliver on its promise of democratizing access to sophisticated trading strategies for retail investors.

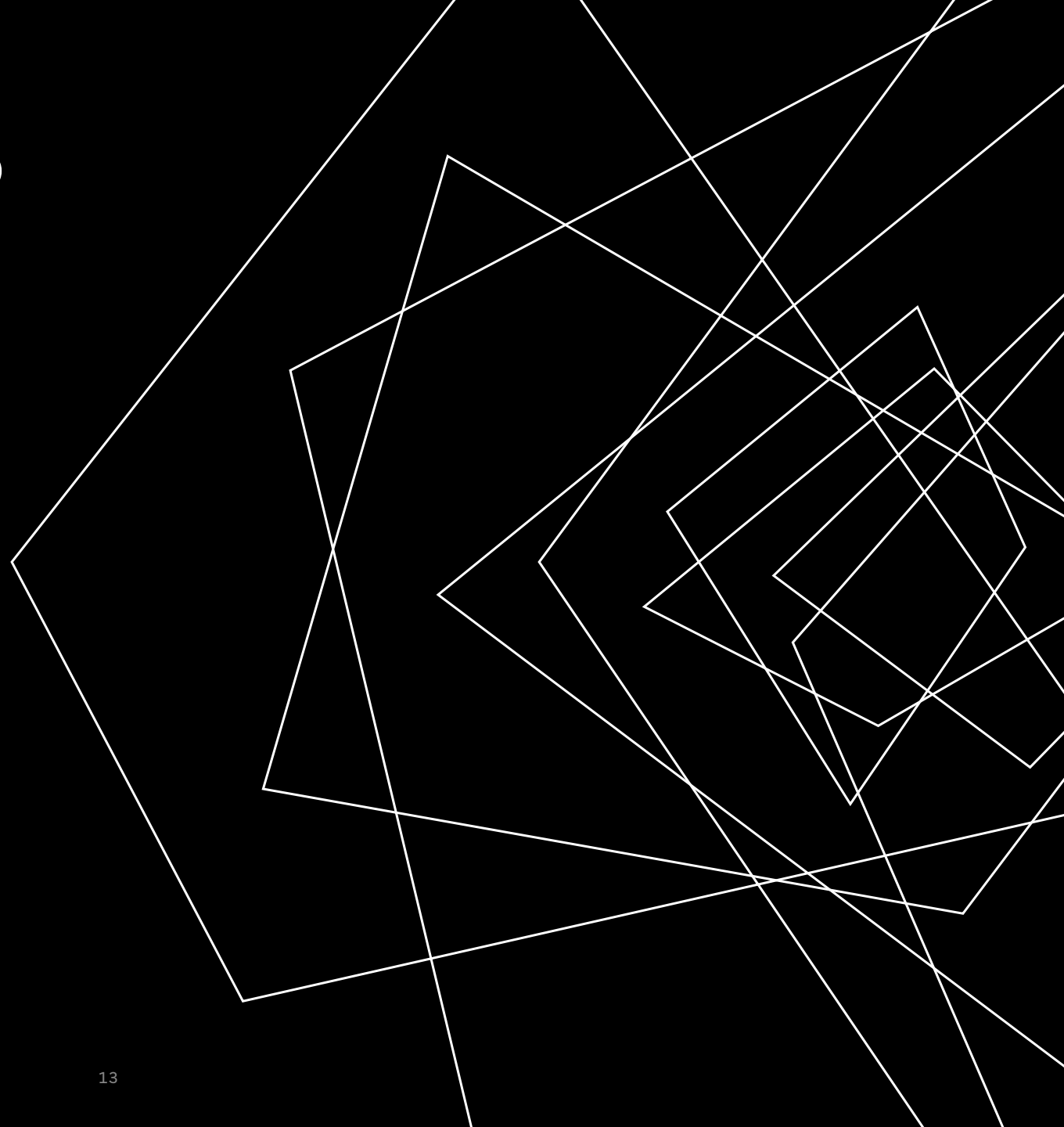
DATA MANAGEMENT AND TRANSFORMATION

01 | Introduction

02 | Dataset Assessment

- Data Usage
- Dataset Transformation
- Dataset Quality Management
- Data Architecture

03 | Conclusion



INCITE

INCITE

INTRODUCTION

OVERVIEW

The purpose of this document is to provide a comprehensive overview of our approach to data management, transformation, and the architectural design we will adopt for our project. It will delve into how we intend to utilize the datasets, the transformations, and preprocessing required, how we plan to manage data quality, and the architecture that will support this whole process. The aim is to present a clear roadmap of how we will convert raw data into actionable insights and inputs for our trading models.

INCITE

DATASET ASSESSMENT

Dataset Usage	Dataset Transformation	Dataset Quality Management	Data Architecture
- We will primarily use the Alpaca Market Data API for historical and real-time stock prices, the Alpaca Trading API for executing trades, and the Alpaca Asset Information API for metadata on assets. Additionally, we will use the Alpaca TA-Lib (Technical Analysis Library).	- Data from different APIs will need to be aggregated at the appropriate level (such as by ticker and date) to ensure a uniform dataset. Any missing or anomalous data points will need to be handled appropriately (filled in, smoothed, or excluded), depending on their nature and impact on the analysis.	- We will implement data validation checks at various stages of the data pipeline to ensure that the data's quality is maintained. This might involve checking for missing values, verifying the consistency of data formats, and checking for any sudden or unexplained jumps in data values.	- Our data architecture will need to be robust and scalable to handle the real-time data flow from Alpha Vantage's APIs and process it efficiently. It might involve a data ingestion pipeline to fetch and store data, a data processing pipeline to aggregate and transform data, and a data serving layer to make processed data available to our trading models.

INCITE

CONCLUSION

KEY TAKEAWAYS

Data management and transformation are critical to the success of the 'Incite' project. By ensuring effective handling, transformation, and structuring of our Alpha Vantage datasets, we can build reliable and efficient algorithmic trading models. Our chosen data architecture will support real-time data flow and processing, ensuring that our trading models can adapt quickly to changing market conditions.

QUESTIONS

INCITE

INCITE

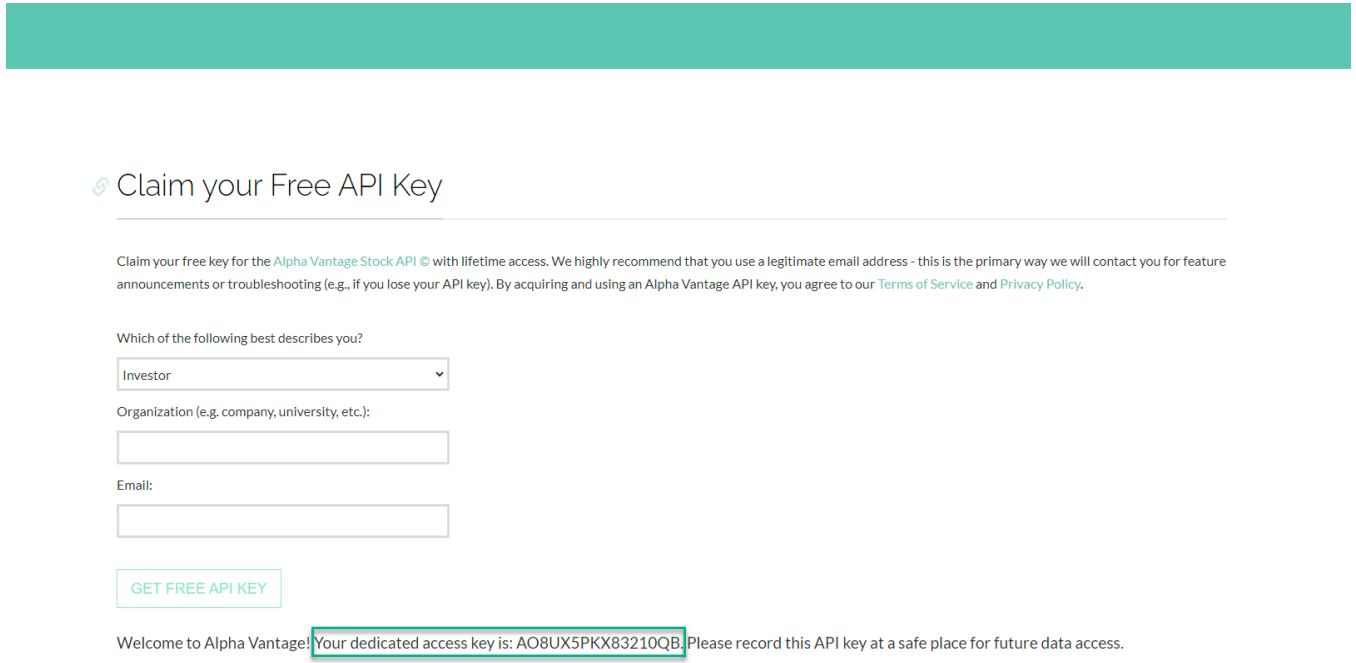
APPENDIX

SOURCES

Type	Trend	Name	Source	Description
Trend 1	Market Volatility	"Historical Volatility"	Investopedia: Historical Volatility	This Investopedia article discusses historical market volatility trends.
Trend 1	Market Volatility	"Stock Market Volatility and Return Analysis: A Systematic Literature Review"	Stock Market Volatility and Return Analysis	This study, published in the National Center for Biotechnology Information, analyzes market volatility.
Trend 2	AI and Machine Learning in Finance	"Understanding How AI is Transforming Financial Trading"	Analytics Insight: Understanding How AI is Transforming Financial Trading	This article from Analytics Insight discusses the transformative role of AI and machine learning in financial trading.
Trend 3	Retail Investor Activity	"The Rise of Newly Empowered Retail Investors"	Deloitte: The Rise of Newly Empowered Retail Investors	This Deloitte report discusses the increasing power and influence of retail investors in the financial markets.
Trend 3	Retail Investor Activity	"Retail Investors in 2023: Performance, Trends and Challenges"	Medium: Retail Investors in 2023: Performance, Trends and Challenges	This Medium article by Nauris Krūmiņš delves into the performance, trends, and challenges retail investors may face in 2023.

APPENDIX

ALPHA VANTAGE ACCESS CODE



APPENDIX

ALPHA VANTAGE ACCESS CODE

ValueError: Thank you for using Alpha Vantage! Our standard API call frequency is 5 calls per minute and 500 calls per day. Please visit https://www.alphavantage.co/premium/ if you would like to target a higher API call frequency.

Define functions to fetch latest price and volume information for a stock

```
def quote_endpoint(symbol):
    data, _ = ts.get_quote_endpoint(symbol)
    return data

def search_endpoint(keywords):
    from alpha_vantage.timeseries import TimeSeries
    ts = TimeSeries(key=api_key)
    data, _ = ts.get_symbol_search(keywords)
    return data
```

```
plot_stock_data('TSLA')  # plot stock data for Apple Inc.
print(quote_endpoint('TSLA'))  # print latest price and volume info for Apple Inc.
print(search_endpoint('TSLA'))  # search for symbols matching 'apple'
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_24584\647870858.py in <module>
----> 1 plot_stock_data('TSLA')  # plot stock data for Apple Inc.
      2 print(quote_endpoint('TSLA'))  # print latest price and volume info for Apple Inc.
      3 print(search_endpoint('TSLA'))  # search for symbols matching 'apple'

~\AppData\Local\Temp\ipykernel_24584\2858298256.py in plot_stock_data(symbol)
      3 def plot_stock_data(symbol):
      4     # Fetch intraday, weekly, and monthly time series data
----> 5     intraday_data, _ = ts.get_intraday(symbol=symbol, interval='60min', outputsize='full')
      6     weekly_data, _ = ts.get_weekly(symbol=symbol)
      7     monthly_data, _ = ts.get_monthly(symbol=symbol)

~\Anaconda3\lib\site-packages\alpha_vantage\alphavantage.py in _format_wrapper(self, *args, **kwargs)
    216         @wraps(func)
    217         def _format_wrapper(self, *args, **kwargs):
--> 218             call_response, data_key, meta_data_key = func(
    219                 self, *args, **kwargs)
    220             if 'json' in self.output_format.lower() or 'pandas' \

~\Anaconda3\lib\site-packages\alpha_vantage\alphavantage.py in _call_wrapper(self, *args, **kwargs)
    158             else:
    159                 url = '{}{}'.format(url, apikey_parameter)
--> 160             return self._handle_api_call(url), data_key, meta_data_key
    161         return _call_wrapper
    162

~\Anaconda3\lib\site-packages\alpha_vantage\alphavantage.py in _handle_api_call(self, url)
    361                 raise ValueError(json_response["Information"])
    362             elif "Note" in json_response and self.treat_info_as_error:
--> 363                 raise ValueError(json_response["Note"])
    364             return json_response
    365         else:

ValueError: Thank you for using Alpha Vantage! Our standard API call frequency is 5 calls per minute and 500 calls per day.
Please visit https://www.alphavantage.co/premium/ if you would like to target a higher API call frequency.
```

APPENDIX

INSTALL REQUIRED PACKAGES

Insert

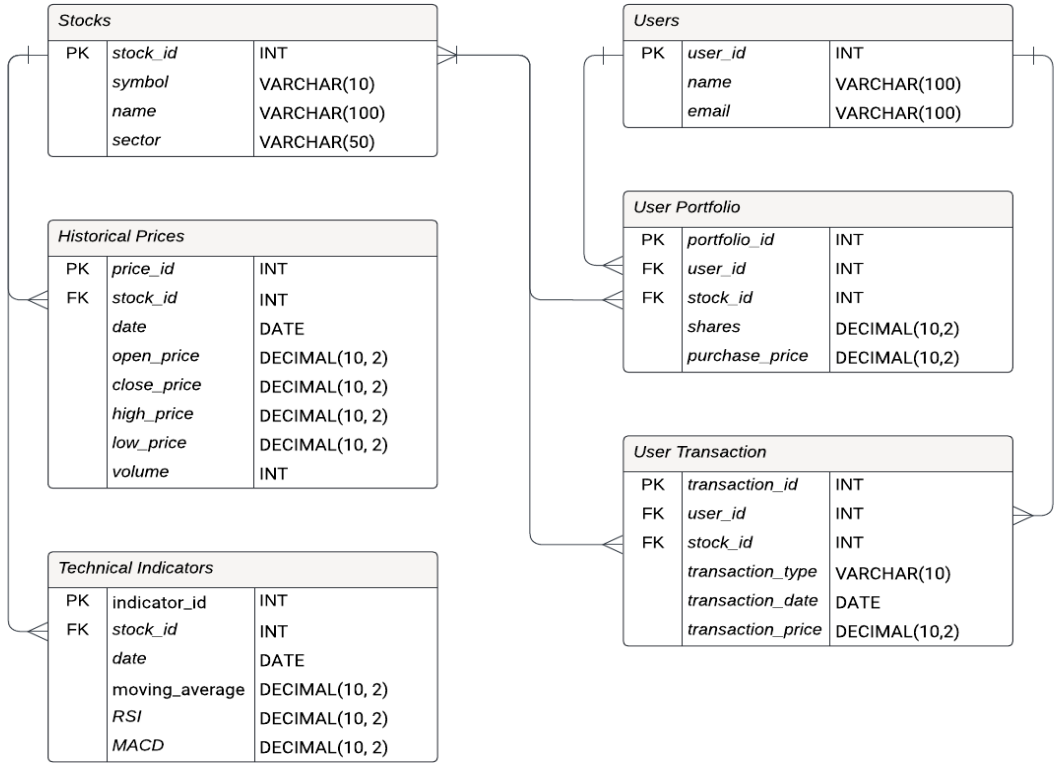
```
1 pip install alpha_vantage pandas matplotlib
```

```
Collecting alpha_vantage
  Downloading alpha_vantage-2.3.1-py3-none-any.whl (31 kB)
Requirement already satisfied: pandas in c:\users\steand\anaconda3\lib\site-packages (1.4.4)
Requirement already satisfied: matplotlib in c:\users\steand\anaconda3\lib\site-packages (3.6.2)
Requirement already satisfied: requests in c:\users\steand\anaconda3\lib\site-packages (from alpha_vantage) (2.28.1)
Requirement already satisfied: aiohttp in c:\users\steand\anaconda3\lib\site-packages (from alpha_vantage) (3.8.3)
Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\steand\anaconda3\lib\site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\steand\anaconda3\lib\site-packages (from pandas) (2022.7)
Requirement already satisfied: numpy>=1.18.5 in c:\users\steand\anaconda3\lib\site-packages (from pandas) (1.23.5)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (4.25.0)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (3.0.9)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (1.0.5)
Requirement already satisfied: pillow>=6.2.0 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (9.3.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (22.0)
Requirement already satisfied: cycler>=0.10 in c:\users\steand\anaconda3\lib\site-packages (from matplotlib) (0.11.0)
Requirement already satisfied: six>=1.5 in c:\users\steand\anaconda3\lib\site-packages (from python-dateutil>=2.8.1->panda
s) (1.16.0)
Requirement already satisfied: yarl<2.0,>=1.0 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->alpha_vantage)
(1.8.2)
Requirement already satisfied: charset-normalizer<3.0,>=2.0 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->a
lpha_vantage) (2.0.4)
Requirement already satisfied: aiosignal>=1.1.2 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->alpha_vantag
e) (1.3.1)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->al
pha_vantage) (4.0.2)
Requirement already satisfied: multidict<7.0,>=4.5 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->alpha_vant
age) (6.0.4)
Requirement already satisfied: frozenlist>=1.1.1 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->alpha_vantag
e) (1.3.3)
Requirement already satisfied: attrs>=17.3.0 in c:\users\steand\anaconda3\lib\site-packages (from aiohttp->alpha_vantage)
(22.1.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\steand\anaconda3\lib\site-packages (from requests->alpha_vant
age) (2022.12.7)
Requirement already satisfied: idna<4,>=2.5 in c:\users\steand\anaconda3\lib\site-packages (from requests->alpha_vantage)
(3.4)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\steand\anaconda3\lib\site-packages (from requests->alpha_v
antage) (1.26.14)
Installing collected packages: alpha_vantage
Successfully installed alpha_vantage-2.3.1
Note: you may need to restart the kernel to use updated packages.
```

INCITE

RELATIONAL DATA MODEL

DATA GAPS



RELATIONSHIP SUMMARY

- Stocks to Historical Prices: One-to-Many (Each stock can have multiple historical price records).
- Stocks to Technical Indicators: One-to-Many (Each stock can have multiple technical indicator records).
- Users to User Portfolio: One-to-Many (Each user can have multiple portfolios).
- Stocks to User Portfolio: Many-to-Many (Many stocks can belong to multiple user portfolios).
- Users to User Transactions: One-to-Many (Each user can have multiple transactions).
- Stocks to User Transactions: Many-to-Many (Many stocks can be involved in multiple user transactions).