**PRD: Data Pipeline**

**Product Engineering Project Capstone**

*by Nick, Amrita, Jasmine, Alejandro*

**Change history:**

* Document created 2/19/2024

**Overview:**

* By using semi-structured data, specifically about daily changes in Amazon stock prices, we will create a data pipeline, utilizing extraction, transforming, loading and analysis to achieve clean, structured data as well as insightful analysis.

**Objectives:**

* Find unstructured data and write a python script to pull data from the source
* Utilize pandas in python to clean and transform the data
* Utilize data normalization strategies for efficient data querying.
* Create a dataframe.
* Load the cleaned data into local storage such as a local database
* Analyze the transformed data to generate meaningful inferential and descriptive statistical insights

**Success Metrics:**

* Have a working data pipeline which updates in real time.
* Clear and efficient data analysis.
* Working data manipulation showcasing vital descriptive & inferential data.

**Messaging:**

* The deliverable of this product will be a MySQL database with transformed and cleaned data as well as statistical insights.

**Timeline and Release Planning:**

* Release Date 2/22/24

**Personas/Scenarios:**

* Alejandro: wants to retire early, and so is trying to come up with an investing strategy that will allow him to buy a lot land and go off the grid
* Jasmine: a finance student wants to wow her professors by providing real time stock trading data.
* Amrita: a business owner who invested a chunk of her earnings into Amazon after learning about AWS, she wants to make sure her money is being invested wisely and getting the returns she expected.
* Nick: wants to get into investing into stocks as a side hustle, so he can use this data to view trends and patterns

**Features In:**

*[M] denotes minimum viable experience requirement for Amazon Stock Data ETL*

* Statistical insights [M]
* MySQL Database [M]

**Features Out:**

* Backup functionality

**Designs:**

* Documentation on data extraction (code and a data sample)
* Report on data transformation, outline rules, challenges and solutions
* Documentation on schema design and loading strategies
* Analytical report on data after transforming and storage into database
* Presentation slides

**Open Issues:**

* None at the moment