



COMPUTER SCIENCE AND DATA ANALYTICS

Data pipeline monitoring and alerting.

Report 2

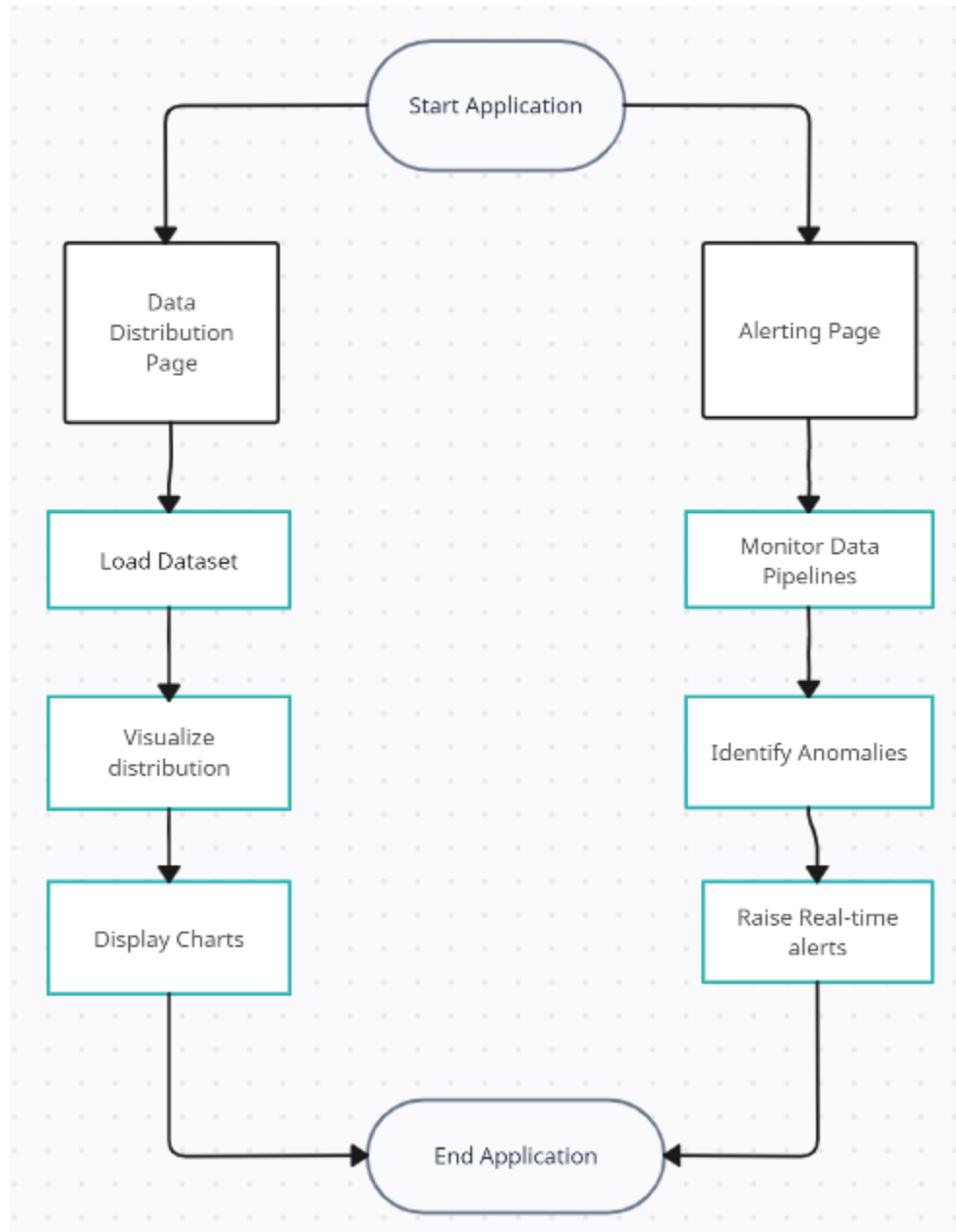
Student: Sokrat Bashirov

GWID: G26315644.

Research strategy:

I am going to use a quantitative research approach in my research topic, "Data Pipeline Monitoring and Alerting", in which the aim is to design and implement a web application using Flask framework. The application will allow banks and industry companies to monitor their data pipelines and receive real-time alerts for anomaly detection. Additionally, the application will provide a separate page to visualize the data distribution and performance metrics of the monitored pipelines.

Flow diagram:



Data collection:

A data collection strategy is not suitable for the topic of my project. This research's main goal is to build and implement a web application applying the Flask framework, that will track and alert users to data pipeline performance indicators like throughput, latency, and error rates.

As the research primarily involves building an app, the dataset used for testing and validation is obtained from secondary sources. More precisely, a dataset from the banking industry was used, which had previously been gathered and used to create a model to forecast client attrition.

Data cleansing approaches:

As we work with applications, the dataset being used to test the program is being intentionally "spoiled" or carefully altered by hand to trigger certain problems in order to evaluate the application's problem identification abilities.