**A logo of a company

Description automatically generatedA logo of a company

Description automatically generated**

REPORT

**Data Acquisition**

In the initial stage of our project, our primary objective was to establish an efficient mechanism for obtaining the necessary data. To achieve this, we leveraged the capabilities of Python, an immensely versatile programming language. Python's involvement was particularly prominent within the Jupyter Notebook environment, a powerful tool widely acclaimed for its interactive and collaborative features. The Jupyter Notebook played a crucial role in our data workflow by facilitating seamless access to a dataset hosted on Github via its URL. This dataset was subsequently stored in a MongoDB database, providing a robust and organized repository for our project's data. The utilization of Jupyter Notebook not only streamlined the retrieval process from an external source but also set the foundation for efficient data exploration and analysis.

## Data Storage

## To establish a robust data storage infrastructure, we initiated the creation of a MongoDB database named 'super\_store,' featuring a collection named 'sales\_details.' To seamlessly integrate this database with the Jupyter Notebook environment, I obtained a connection link, fostering a direct and efficient link between the notebook and the database. The dataset, integral to our project, found its secure repository within MongoDB, ensuring proper organization and accessibility. Furthermore, recognizing the need for enhanced storage capabilities and adherence to industry standards, the project strategically harnessed the robust features of the Microsoft Azure server for cloud-based storage. This strategic decision not only fortified the security of our stored data but also positioned our project within the realm of scalable and industry-compliant cloud solutions, marking a critical stride toward reliable and future-ready data storage infrastructure.

## Data Exploration and Analysis

Conducting a comprehensive exploration of the dataset served as a foundational step to gain insights into its intricacies. This involved a meticulous examination of key aspects, including the number of variables, the volume of records, the nature of variables, and their respective data types. The exploration process delved into unraveling intricate relationships existing between these variables. By systematically dissecting these components, we aimed to establish a nuanced understanding of the dataset's structure, paving the way for informed decision-making in subsequent analytical phases. Also we explored the dataset to identify and address potential data issues. This encompassed thorough checks for duplications, outliers, and missing values within the dataset.

## Data / Web Visualization

We harnessed the capabilities of Flask, a robust web application framework, to craft a dynamic and interactive data visualization platform. We were able to seamlessly integrate our data visualizations into a web environment, thereby enhancing accessibility and user interaction. Flask's flexibility allowed us to create a tailored and responsive interface that catered to the diverse needs of our audience. The resulting web application, hosted on a Python server, played a pivotal role in democratizing access to our project's findings. Beyond just presenting information, it provided a user-friendly interface that facilitated effortless navigation through the visualized data.