**Project Topic:**

* **Evaluating the Impact of Marketing Strategies through A/B Testing**

**Project Overview:**

* This project aims to assess the effectiveness of different marketing strategies by analyzing data from an A/B test. The A/B test involves two groups: a control group and a test group. The control group follows the existing marketing strategy, while the test group experiences a new or modified strategy. The datasets provided contain various key performance indicators (KPIs) such as USD Spent, Number of Impressions, Reach, Number of Website Clicks, Number of Searches, Number of Viewed Content, Number of Added to Cart, and Number of Purchase. The goal is to determine which strategy is more effective in driving desired user behaviors and optimizing marketing spend.

**Implementation Steps:**

1. **Data Collection:**
   * This data collection is from Kaggle with a 10.0 Usability
   * <https://www.kaggle.com/datasets/amirmotefaker/ab-testing-dataset>
2. **Data Ingestion:**
   * Use Azure Data Factory to orchestrate the data pipeline, ingesting data from cloud storage into Azure Storage Blob.
   * Automate data extraction and load the datasets into Databricks for processing.
3. **Data Cleaning and Preprocessing:**
   * Utilize Databricks with PySpark to process and clean the data from both control and test groups.
   * Write PySpark code to clean the data, handle missing values, remove duplicates, and standardize formats.
4. **Data Analysis and Visualization:**
   * Use Databricks notebooks for Exploratory Data Analysis (EDA) with libraries like Pandas, NumPy, Matplotlib, and Plotly.
   * Create visualizations to explore the distribution of data, identify trends, and spot anomalies.
5. **Hypothesis Formulation:**
   * Develop hypotheses to test the new marketing strategy's effectiveness, such as increasing sales, followers, or traffic.
   * Example Hypothesis: The test group has a higher average number of purchases compared to the control group.
6. **Data Storage and Management:**
   * Use cloud platforms (Azure) for deployment and scaling.
     1. Azure Storage Accounts > Containers
     2. Azure Databricks > Create Notebooks > Create Cluster > Multi-Node > Autoscaling
     3. Azure Data Factory > Pipeline > Linked Services > Connect Notebooks
   * Store processed data and analysis results in Azure Storage Blob, ensuring data is secure and accessible for future use.
   * Use version control Git Configuration to manage different versions of datasets and models.
7. **Automation and Scheduling:**
   * Automate the entire workflow using Azure Data Factory. Set up schedules for regular data ingestion, processing, and analysis.
   * Monitor pipeline performance and handle any exceptions or errors.
8. **Marketing Strategy Collaboration:**
   * Work with marketing and customer support teams to design and implement strategies.
   * Monitor the effectiveness and adjust based on feedback.

**Expected Outcomes:**

* Identification of the most effective marketing strategy to achieve business goals.
* Increased customer traffic leading to higher conversion rates.
* Enhanced revenue and profitability through targeted market strategies.

### **Next Steps:**

* Create dashboards and reports using Databricks notebooks or integrate with tools like Power BI or Tableau for real-time visualization and reporting.
* Share insights and recommendations with stakeholders through interactive dashboards.

**Tools and Technologies:**

* **Programming Languages:** PySpark, Python, SQL
* **Libraries:** Pandas, NumPy, Scikit-Learn, Matplotlib, Plotly, Azure Storage Blob
* **Cloud Platforms:** AWS, Azure
* **Data Platforms and Tools:** DataBricks, Azure Data Factory, Azure Storage Blob & Gen2