*Gasser Ahmed*

[*gasser18@vt.edu*](mailto:gasser18@vt.edu)

*ECE 5984, Project Milestone*

**KickStats: Visualizing Live Football Scores**

**What have been done so far?**

1. **GitHub:** Created a GitHub repo for the project (*https://github.com/gasserahmed/kickstats*)
2. **API-FOOTBALL:** Created an app (KickStats) under Rapid API where I generated an API key to be use for the app's endpoints calls where I'm using *“V3 - Fixtures in progress (LiveScore)”* endpoint to get the live game scores.
3. **Data Ingestion** **& Stream Ingestion (Extract and Load):** Implemented the stream ingestion logic and finished the ingestion process where I pushed data successfully to *s3://ece5984-bucket-gasser18/Project/kickstats-stream* via Kafka consumer and producer.
4. **Data Storage (S3 Bucket):** Created a separate directory for the project at my S3 bucket ([*https://s3.console.aws.amazon.com/s3/buckets/ece5984-bucket-gasser18*](https://s3.console.aws.amazon.com/s3/buckets/ece5984-bucket-gasser18)) that contains LiveScore data for 10/15/2023

**What pipeline, project, and dataset are being used?**

For this project, we will use the following Stream-Visualization pipeline for the [API-FOOTBALL](https://www.api-football.com/)’ s LiveScore dataset:

1. **Data Ingestion:** We will retrieve real-time football match data from the API-Football API.
2. **Stream Ingestion (Extract and Load):** Using Apache Kafka, we will ingest and stream the live data to ensure we have the latest scores and updates.
3. **Data Storage:** The ingested data will be stored in Amazon S3, forming a data lake that allows for scalable and cost-effective storage.
4. **Data Transformation:** Pandas, a powerful Python library for data manipulation, will be used to transform the raw data into a structured format suitable for analysis.
5. **Data Warehousing:** The transformed data will be further stored in an Amazon S3 bucket, serving as our data warehouse for historical data analysis.
6. **Relational Database:** We will use Amazon RDS or MySQL to maintain a relational database for structured data storage and querying.
7. **Data Analytics:** Tableau will be employed to create interactive dashboards and visualizations, enabling users to access live football match scores with ease.

**Does data need data cleaning or preprocessing?**

The data retrieved from the "KickStats" app is relatively clean, but it may require minor preprocessing to structure it for visualization. This may include organizing scores, match details, and related statistics.

**Will the project perform Exploratory Data Analysis? Which methods will be used?**

N/A

Exploratory Data Analysis (EDA) will be performed to gain insights into the dataset. EDA techniques such as summary statistics, data distribution visualizations, and trend analysis will be employed.

**What information about data provenance have been listed? Answer the characteristic data provenance questions addressed in Module 5**

N/A

**Are there any untested assumptions or other reasons that would prevent me from completing the project?**

N/A

As of this milestone, there are no untested assumptions or significant barriers that would prevent project completion. The data stream ingestion process has been successfully implemented.

**What are the next steps?**

**Data Transformation:** Use Pandas to transform the raw data into a structured format suitable for analysis.

**What is left to do?**

1. Data Transformation
2. Data Warehousing
3. Relational Database
4. Data Analytics
5. Visualization