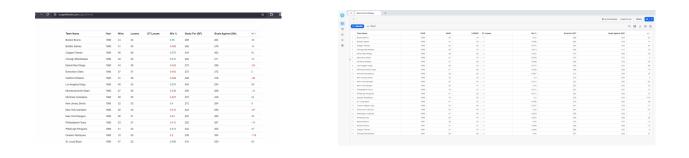# Automated Data Pipeline for Hockey Team Analytics

Technologies: Python (BeautifulSoup, pandas), PySpark, Snowflake, Web Scraping

- Developed an end-to-end automated data pipeline to scrape, process, and analyze hockey team data, leveraging Python for web scraping (BeautifulSoup) and data manipulation (pandas).
- Engineered a scalable storage solution in Snowflake, utilizing PySpark for efficient handling and loading of large datasets, enhancing the data's accessibility for real-time analytics.
- Implemented dynamic database schema creation, automating the mapping of data to Snowflake table structures, significantly reducing manual schema alignment efforts.
- Facilitated advanced data analytics by providing a structured, query-ready dataset in Snowflake, supporting stakeholders in making informed decisions based on up-to-date hockey team insights.
- Achieved a streamlined workflow from data collection to analysis, enabling timely and insightful analytics that support strategy formulation for hockey teams and enthusiasts.

Source Data And Stored Data

Web Scrape ETL Process Design