**Formula 1 (2019) Analysis**

**Overview**

Formula One (more commonly known as Formula 1 or F1) is the highest class of international racing for open-wheel single-seater formula racing cars sanctioned by the Fédération International de l'Automobile (FIA). The FIA Formula One World Championship has been one of the premier forms of racing around the world since its inaugural season in 1950. The word formula in the name refers to the set of rules to which all participants' cars must conform. A Formula One season consists of a series of races, known as Grands Prix. Grands Prix take place in multiple countries and continents around the world on either purpose-built circuits or closed public roads.

A points system is used at Grands Prix to determine two annual World Championships: one for the drivers, and one for the constructors (the teams). Each driver must hold a valid Super License, the highest class of racing license issued by the FIA, and the races must be held on tracks graded "1", the highest grade-rating issued by the FIA for tracks.

We are going to perform analysis on the formula 1 data for 2019. Data is going to be collected from different sources and then combined to make a common data. From where it will be used for analysis.

**Terminologies**

**What is ETL?**

In computing, extract, transform, load (ETL) is a three-phase process where data is extracted, transformed (cleaned, sanitized, scrubbed) and loaded into an output data container. The data can be collated from one or more sources, and it can also be outputted to one or more destinations. ETL processing is typically executed using software applications, but it can also be done manually by system operators. ETL software typically automates the entire process and can be run manually or on reoccurring schedules either as single jobs or aggregated into a batch of jobs.

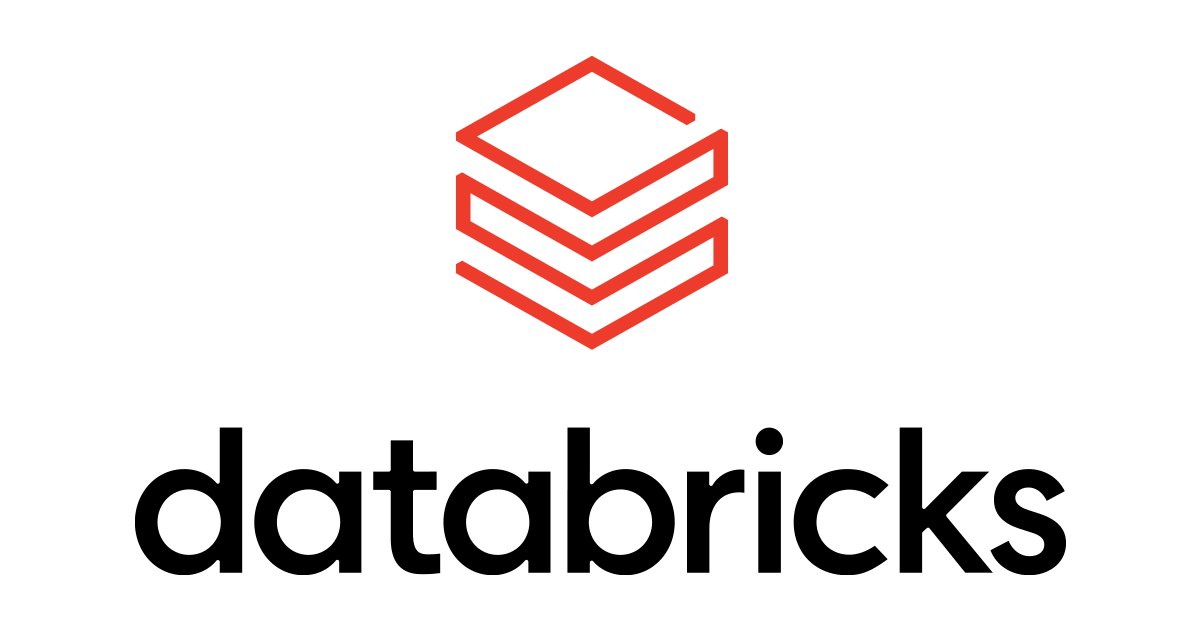
**What is Data Engineering?**

Data engineering refers to the building of systems to enable the collection and usage of data. This data is usually used to enable subsequent analysis and data science, which often involves machine learning. Making the data usable usually involves substantial compute and storage, as well as data processing and cleaning.

**What is Data Analysis?**

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.[1] Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains.[2] In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

**Tools Used**

**What is Databricks?**

The Databricks Lakehouse Platform provides a unified set of tools for building, deploying, sharing, and maintaining enterprise-grade data solutions at scale. Databricks integrates with cloud storage and security in your cloud account and manages and deploys cloud infrastructure on your behalf.

**What is Azure SQL?**

Microsoft Azure SQL Database (formerly SQL Azure, SQL Server Data Services, SQL Services, and Windows Azure SQL Database) is a managed cloud database (PaaS) provided as part of Microsoft Azure.

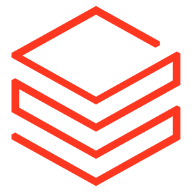
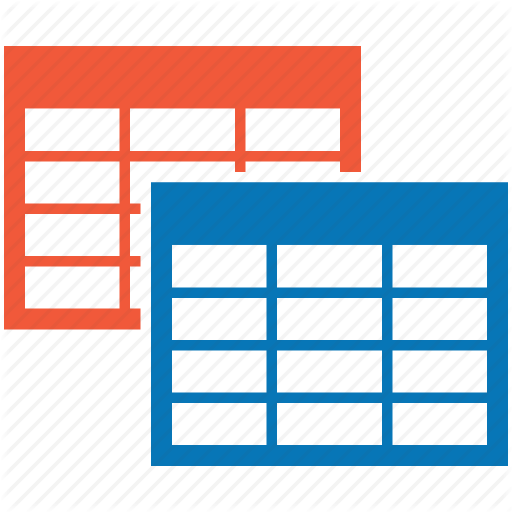
A cloud database is a database that runs on a cloud computing platform, and access to it is provided as a service. Managed database services take care of scalability, backup, and high availability of the database. Azure SQL Database is a managed database service which differs from AWS RDS, which is a container service.

**What is Power BI?**

Power BI is a technology-driven business intelligence tool provided by Microsoft for analyzing and visualizing raw data to present actionable information. It combines business analytics, data visualization, and best practices that help an organization to make data-driven decisions.

**Proposed Methodology**







Databricks

**Report**

**Load**

**Transformation**

**Data**

**Ingestion**

* **Data Source Description**