**The Life Cycle of Data Science**

* **Six Phases of Data Science Lifecycle**:

The data science lifecycle consists of six phases:

Explore Business ---> Data Discovery ---> Data Preparation ---> Exploratory Analysis ---> Model Design ---> Model Development ---> Model Deploy.

Each phase is crucial for the success of a data science project.

* **Exploring the Business**:

This initial phase involves collaboration between the businessperson and the data scientist to understand the main objectives, desired outcomes, and pain points of the business, leading to the definition of a clear and meaningful question to be answered.

* **Data Discovery and Preparation**:

After defining the question, the data discovery phase ensures the necessary data is available, followed by data preparation, where data is organized and combined as needed. This step's complexity varies based on the business's data sophistication.

* **Exploratory Analysis and Model Design**:

In the exploratory analysis phase, the relationships, distributions, and correlations within the data are examined. This understanding guides the model design phase, where the mathematical approach to achieve desired results is defined.

* **Model Building and Deployment**:

The model is then built using programming languages like R or Python. After building the model, it is deployed into the business process, allowing the business to utilize it for decision-making, such as optimizing prices for thousands of products across hundreds of stores.

**Descriptive Analytics**

1. **Importance of Descriptive Analytics**: Descriptive analytics, although not typically considered data science, is crucial as a foundation for more advanced analytics. It helps to understand the current situation before diving into predictive, prescriptive, or diagnostic analytics.
2. **Goals:** Looks Trends and Explain Variances.
3. **Practical Example**: For a grocery store owner, descriptive analytics can identify top and low-selling products. This information is used to optimize shelf space and improve revenue.
4. **Dashboards**: Effective descriptive analytics often involve interactive dashboards (e.g., Tableau, Power BI) that allow users to drill down into data, such as revenue by product category, to make informed decisions.
5. **Data Processing Steps**:
   * **Cleaning**: Ensuring data quality by handling inconsistencies and missing values.
   * **Relating**: Linking data from different sources to provide comprehensive insights.
   * **Summarizing**: Aggregating data to provide an overarching view, such as monthly sales totals per product, Area wise product sales etc.
   * **Visualizing**: Creating user-friendly visual representations that accurately convey the data.
6. **Characteristics and Purpose**: Descriptive analytics is defined by its role in answering questions about the current state of the business. It involves extensive data preparation and visualization to support business decision-making.
7. **Historical Data:**
8. Internal Data
9. External Data
10. **Advantages and Dis-advantages:**
11. Simple and easy to understand
12. Easy comparison
13. Widely accepted
14. Lack of Predictability
15. Lack of Handling Huge Volume of Data
16. **Analysis:**
17. Variance Analysis (It’s an Ad hoc Exercise: Which tells the difference between the outcomes from a train model and the actual business scenario. This will help to find out the gaps and take measures to fill it up)
18. Trend Analysis
19. Comparative Analysis

**Diagnostics Analytics:**

* **Definition and Purpose**: Diagnostic analytics is an advanced phase of analysis focused on understanding the reasons behind data patterns, specifically answering "why" certain events happen by identifying drivers and correlations.
* **Example Scenario**: Using the context of grocery store revenue, one can explain how diagnostic analytics helps determine factors influencing revenue changes, such as time of year, commercial airtime, Twitter mentions, location, etc.
* **Correlation Analysis**: Diagnostic analytics involves analyzing correlations between different variables and revenue, using statistical methods to identify high or low correlations (ranging from 0 to 1).
* **Visualization and Insights**: Visual examples show how variables like Twitter mentions and time of year impact sales differently. This helps in visualizing and understanding the data better.
* **Importance in Data Analysis**: Diagnostic analytics builds on descriptive analytics, allowing for deeper data discovery, drilldowns, and identification of key drivers. It sets the stage for further predictive and prescriptive analytics.