**PROJECT ON : - BIGMART SALES DATA SET PREDICTION**

GROUP MEMBERS: **1. JITENDRA (16HCS4423)**

**PROJECT DETAILS:**

Retail is another industry which extensively uses analytics to optimize business processes. Tasks like product placement, inventory management, customized offers, product bundling, etc. are being smartly handled using data science techniques. As the name suggests, this data comprises of transaction records of a sales store. This is a regression problem.

**PROCEDURE**

The steps taken by us are mentioned below:

1. Hypothesis Generation

This is a very pivotal step in the process of analyzing data. This involves understanding the problem and making some hypothesis about what could potentially have a good impact on the outcome. This is done BEFORE looking at the data, and we end up creating a laundry list of the different analysis which we can potentially perform if data is available.

2. Data Exploration

We’ll be performing some basic data exploration here and come up with some inferences about the data. We’ll try to figure out some irregularities and address them in the next section.

3. Data Cleaning

This step typically involves imputing missing values and treating outliers. Though outlier removal is very important in regression techniques, advanced tree based algorithms are impervious to outliers. So I’ll leave it to you to try it out. We’ll focus on the imputation step here, which is a very important step.

4. Feature Engineering

We explored some nuances in the data in the data exploration section. Lets move on to resolving them and making our data ready for analysis. We will also create some new variables using the existing ones in this section.