**DS Case Study/Assignents:**

1. **Forecasting**

Forecast the CocaCola prices of Airlines Passengers data set. Prepare a document for each model explaining how many dummy variables you have created and RMSE value for each model. Finally which model you will use for Forecasting.

Dataset: CocaCola\_Sales\_Rawdata.xlsx

1. **Decision Tree/Random Forest**

About the data: Let’s consider a Company dataset with around 10 variables and 400 records. The attributes are as follows: → Sales -- Unit sales (in thousands) at each location → Competitor Price -- Price charged by competitor at each location → Income -- Community income level (in thousands of dollars) → Advertising -- Local advertising budget for company at each location (in thousands of dollars) → Population -- Population size in region (in thousands) → Price -- Price company charges for car seats at each site → Shelf Location at stores -- A factor with levels Bad, Good and Medium indicating the quality of the shelving location for the car seats at each site → Age -- Average age of the local population → Education -- Education level at each location → Urban -- A factor with levels No and Yes to indicate whether the store is in an urban or rural location → US -- A factor with levels No and Yes to indicate whether the store is in the US or not The company dataset looks like this:

Problem Statement: A cloth manufacturing company is interested to know about the segment or attributes causes high sale. Approach - A decision tree can be built with target variable Sale (we will first convert it in categorical variable) & all other variable will be independent in the analysis.

Data set: Company\_Data.csv