**Case Study: Help Jeff’s Party Planet to clean its data and analyze it.**

**Introduction:**

Welcome to Jeff’s Party Planet data case study! Jeff’s Party Planet is an event management company that has provided us with the data for cleaning and performing analysis for it. I am assuming I am a freelance junior data analyst who has been approached by Jeff’s Party Planet. The data contains an order listing items ordered for an event, the product quantities ordered, suppliers, price, expenditure per piece, revenue, total expenditure, and profit.

To answer the key business questions, I will follow the steps of the data analysis process: ask, prepare, process, analyze, share, and act.

**Scenario:**

I am a Junior Data Analyst doing freelance data analysis tasks for over a year. I have been recently contacted by Jeff’s Party Planet, a startup event management company. They have recently completed a successful event for their customer. They have provided raw data in a .csv file format containing 34 rows and 8 columns for various products and quantities purchased from four suppliers namely Eco-Disposables, Inflatibles Plus, Sparklefest Ltd., and Supply 4U. The management of Jeff’s Party Planet believes that an analysis of their data will provide them with great insights that will be helpful for their future events. They have asked me to focus on expenditure, profitability, and profit margin per supplier. The insights I discover will help guide their supplier selection strategy for future events.

**Suppliers:**

1. Eco-Disposables,
2. Inflatibles Plus,
3. Sparklefest Ltd., and
4. Supply 4U

**The Data Analysis Process**

1. **ASK**

The management of Jeff’s Party Planet has asked me to clean their data for any irregularities in the data (if any), and analyze the data to make high-level recommendations for suppliers selection process based on total expenditure, total profit, and profit margin per supplier.

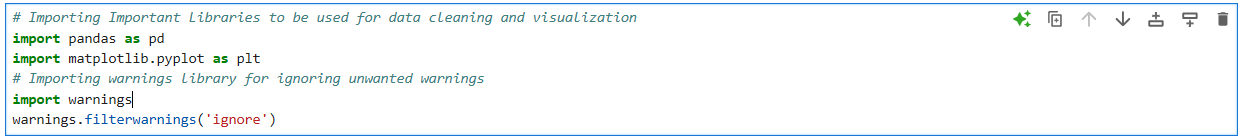
1. **Prepare**

The management of Jeff’s Party Planet has provided me with a dataset in a .csv file format named Jeff\_s\_Party\_Planet.csv through email. I have saved the data in the following location on my laptop f:/Excel Data/ Jeff\_s\_Party\_Planet.csv. They have asked me to clean their data for any irregularities in the data (if any), and analyze the data to make high-level recommendations for suppliers selection process based on total expenditure, total profit, and profit margin per supplier.

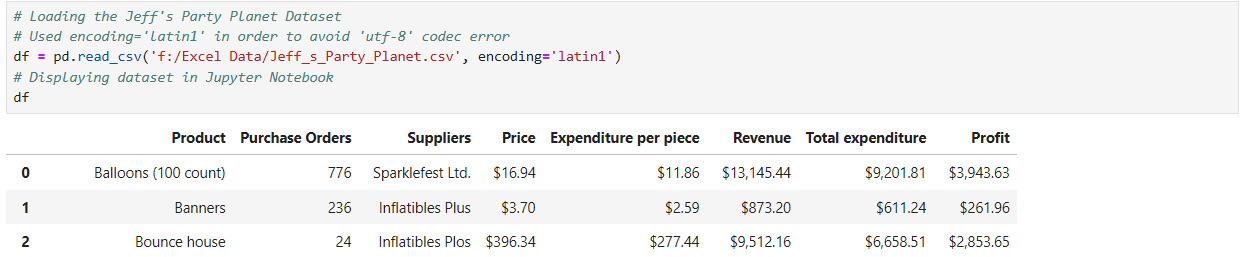
1. **Process**

I have decided to use Python Pandas Library to load, clean, and analyze Jeff’s Party Planet dataset. The steps I have taken to process data are listed below:

1. Open the already installed Jupiter Notebook
2. Install the following important libraries



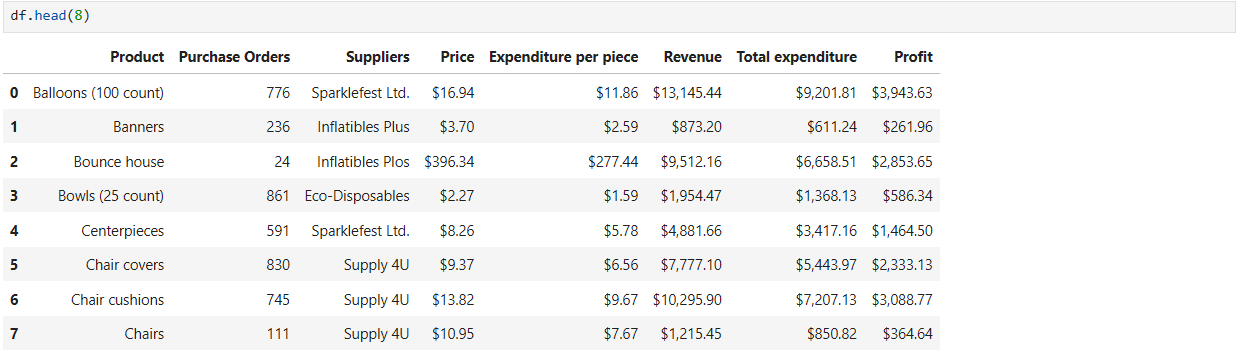
1. Load the data from ‘f:/Excel Data/ Jeff\_s\_Party\_Planet.csv’ and rename the dataframe as df



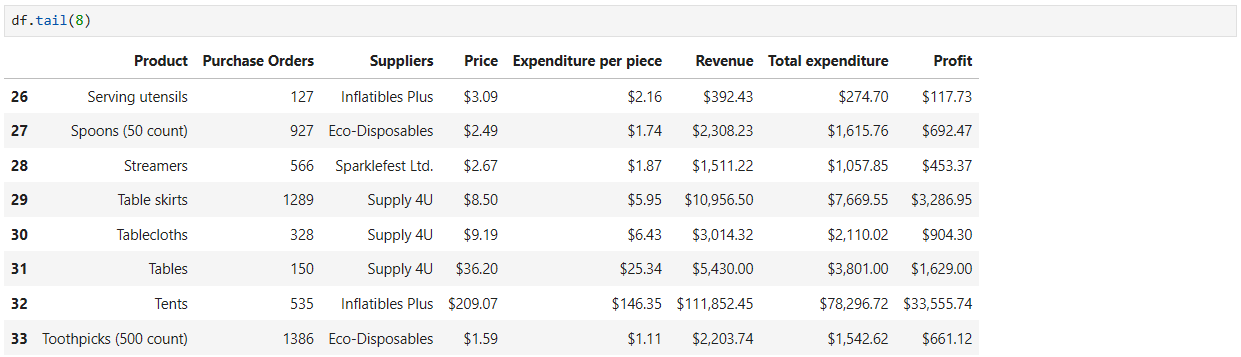
1. **Analyze**

I have used the following methods to analyze dataset:

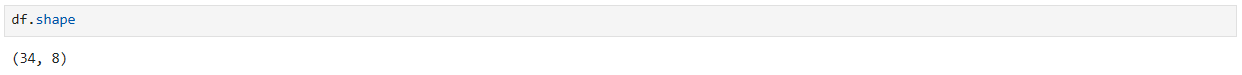
1. Load leading 8 rows to understand the dataset



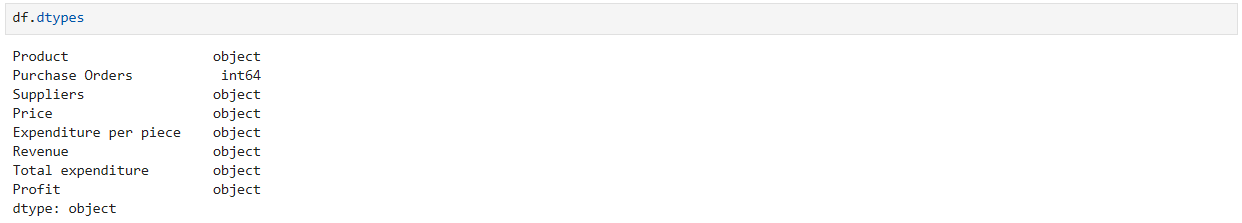
1. Load tailing 8 rows to understand the dataset



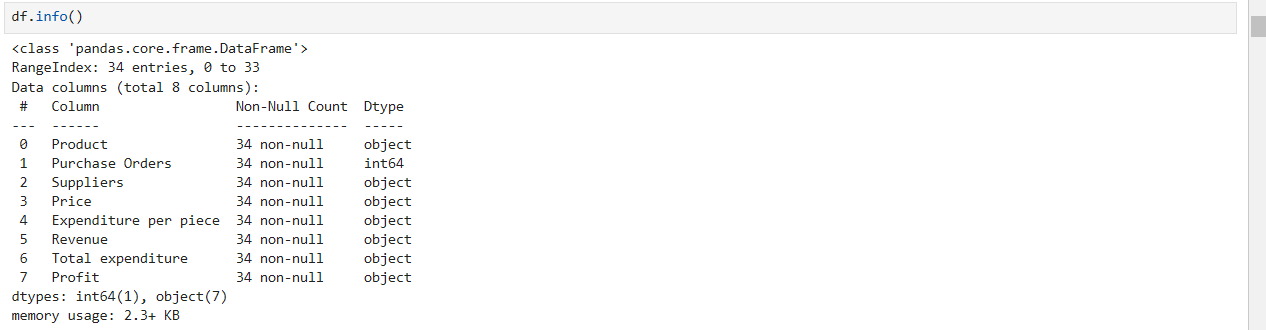
1. Check number of rows and columns in the dataset



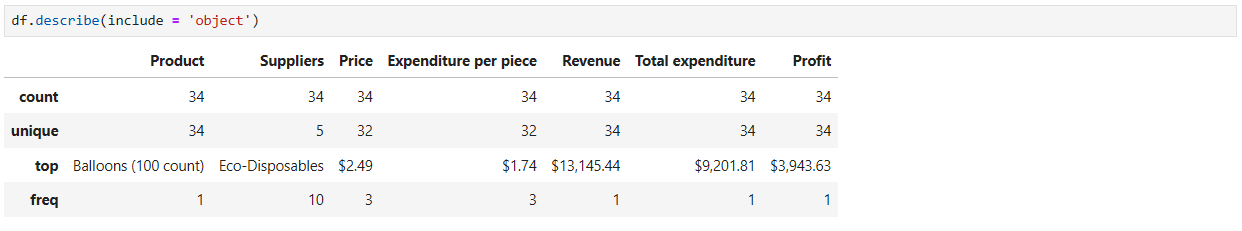
1. Check the data types of the columns



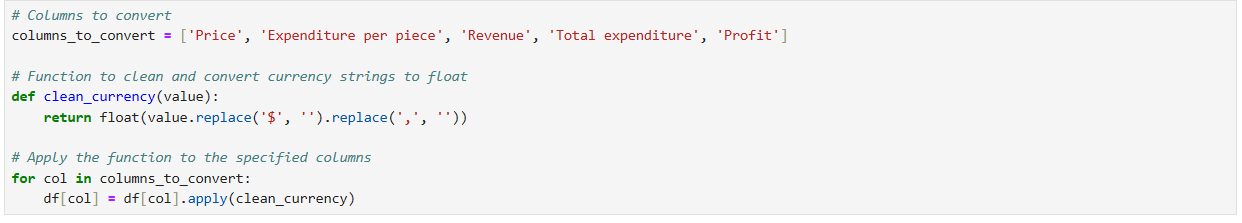
1. Check for datatypes and null values in dataset



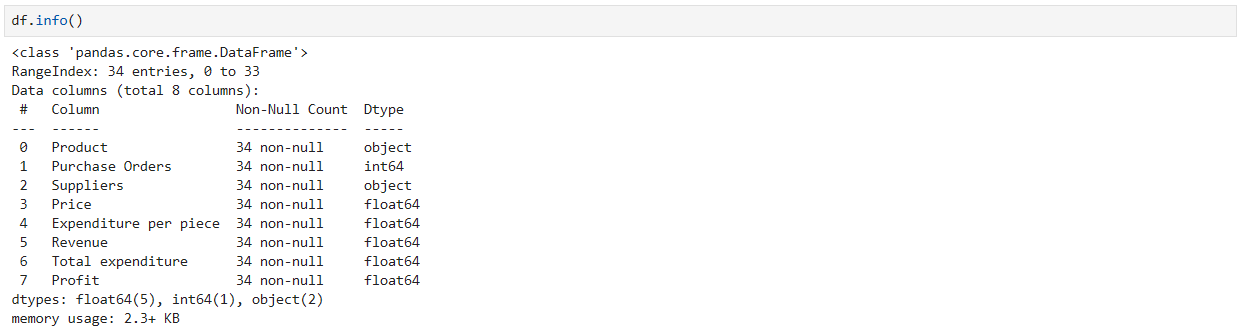
1. Observation: It has been observed through #5 and #6 that the data types in columns Price, Expenditure per piece, Revenue, Total expenditure, and Profit appear as ‘object’ data type instead of float as they are numerical columns instead of strings
2. Let’s further investigate to see why the float data type appears as ‘object’ data type



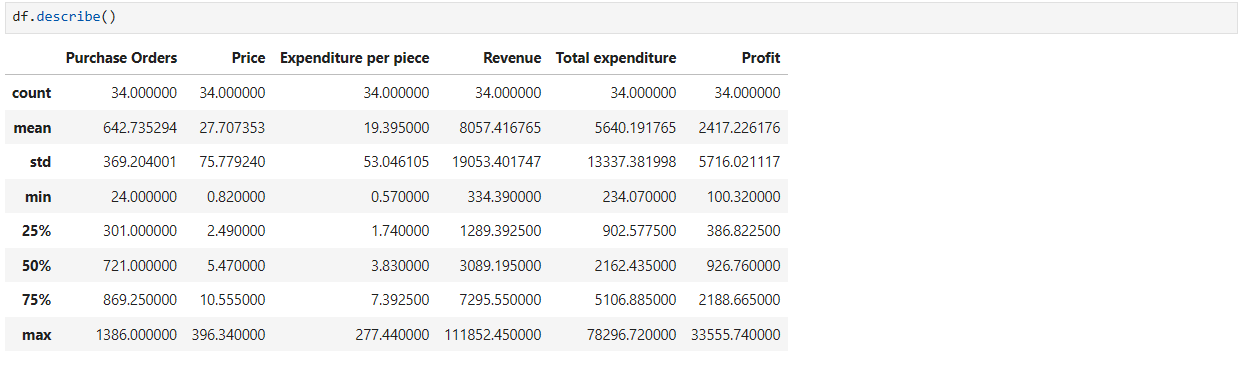
1. Observation: It has been observed above that due to $ sign the pandas interpret the cells as ‘object’ datatype instead of float
2. Let’s convert the object datatype to float



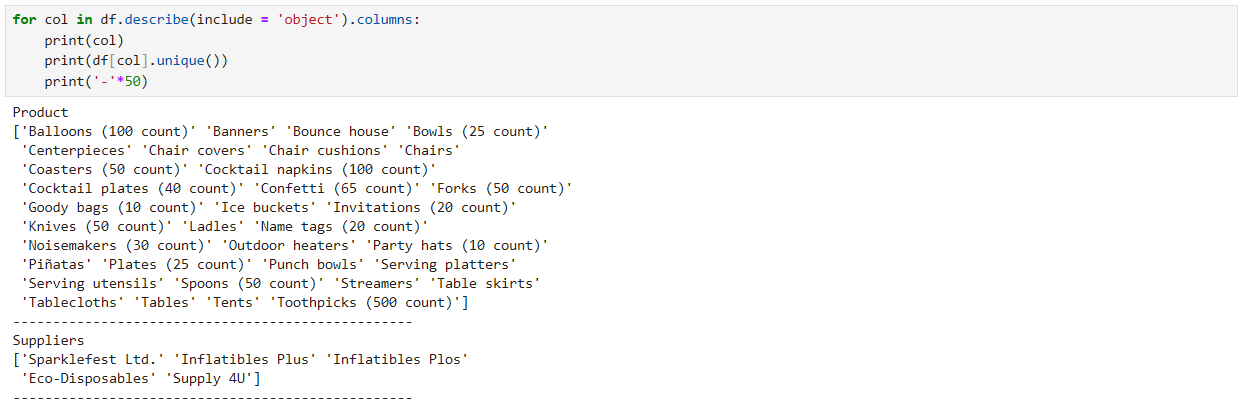
1. Verify if the data type has been changed



1. Use statistical methods to find any outliers in the dataset. From the following, it has been observed that there are no outliers in the dataset.

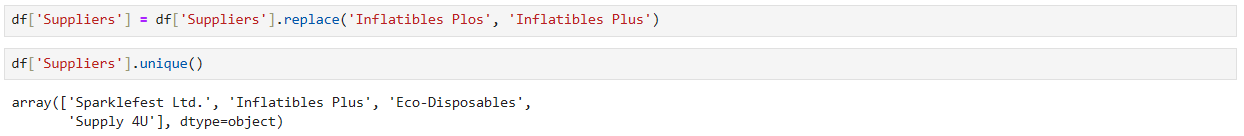


1. Since we are analyzing the dataset regarding suppliers it is important to query and errors in the supplier's column. The data type of the suppliers’ column is object therefore it is important to query columns with object data type

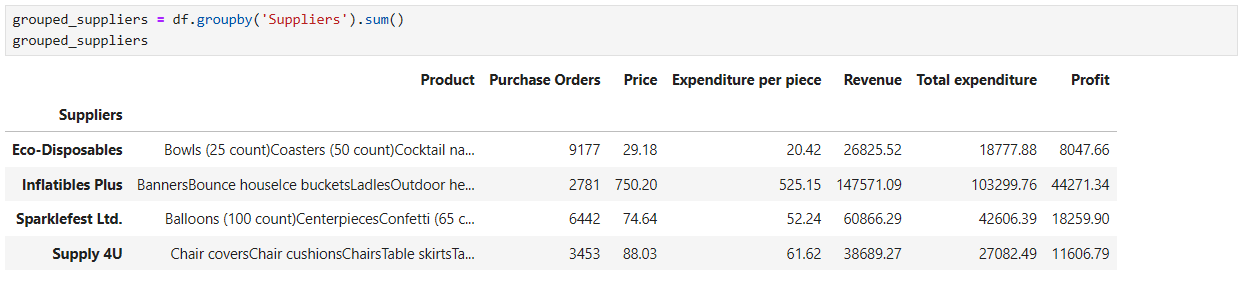


1. Observation: It has been observed there is an error in suppliers column the ‘Inflatibles Plus’ has been misspelled as ‘Inflatibles Plos’.

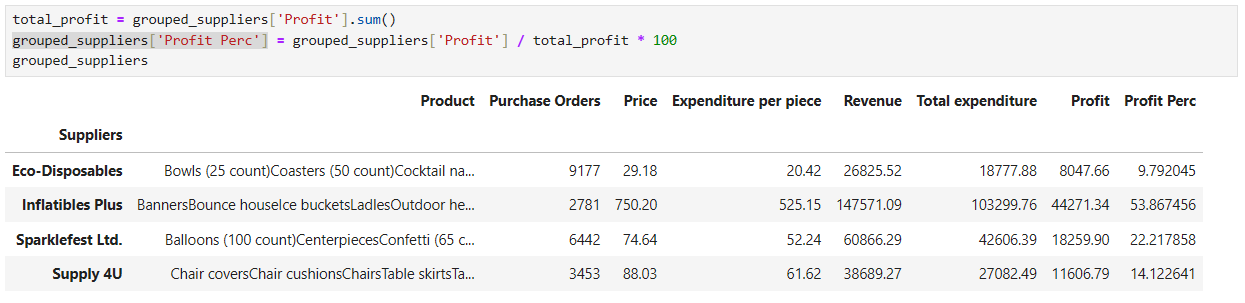
1. Correct the supplier name Inflatibles Plos to Inflatible Plus



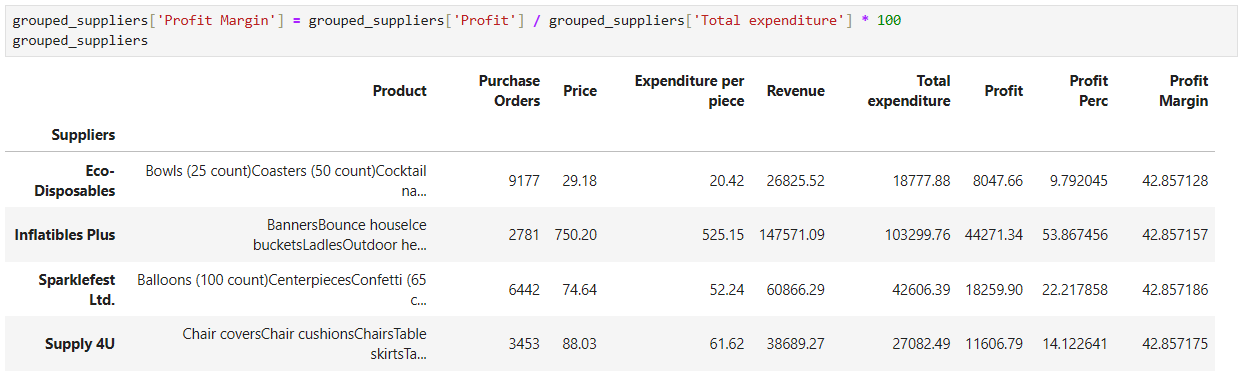
1. Group the data by suppliers to simplify for finding trends



1. Let’s analyze profitability per supplier



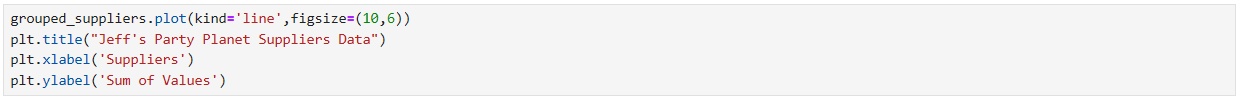
1. Let us analyze profit margin per supplier

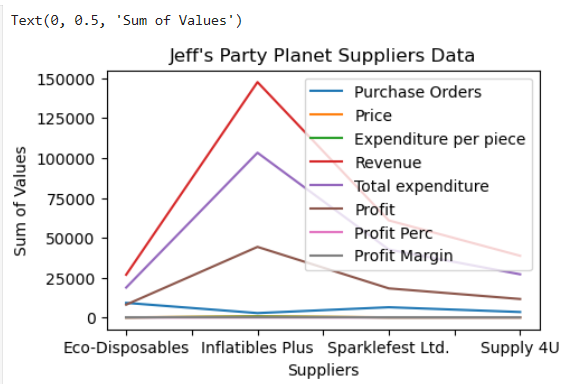


1. **Share**

I have finished analyzing data. Let’s visualize it through visualization in matplotlib library of Python.

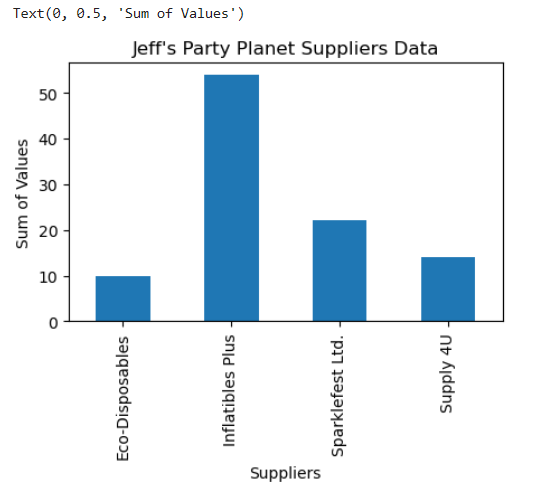
1. Visualize the data by suppliers to find patterns





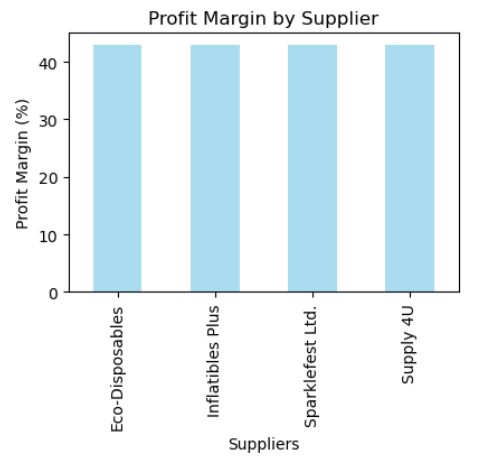
1. Observation: From the above visualization it can be said that Jeff’s Party Planet ordered less from Inflatibles Plus but got high profits.
2. Let’s visualize profitability per supplier





1. Observation: The above visualization shows that Jeff’s Party Planet has drived higher profits Inflatible Plus
2. Let’s visualize profit margin per supplier





1. From the above visualization it is evident that all the suppliers are equally profitable.
2. **Act**