**The University of Texas**

**at Dallas**

**MIS 6309.002**

**Business Data Warehousing**

**Use case development topic:**

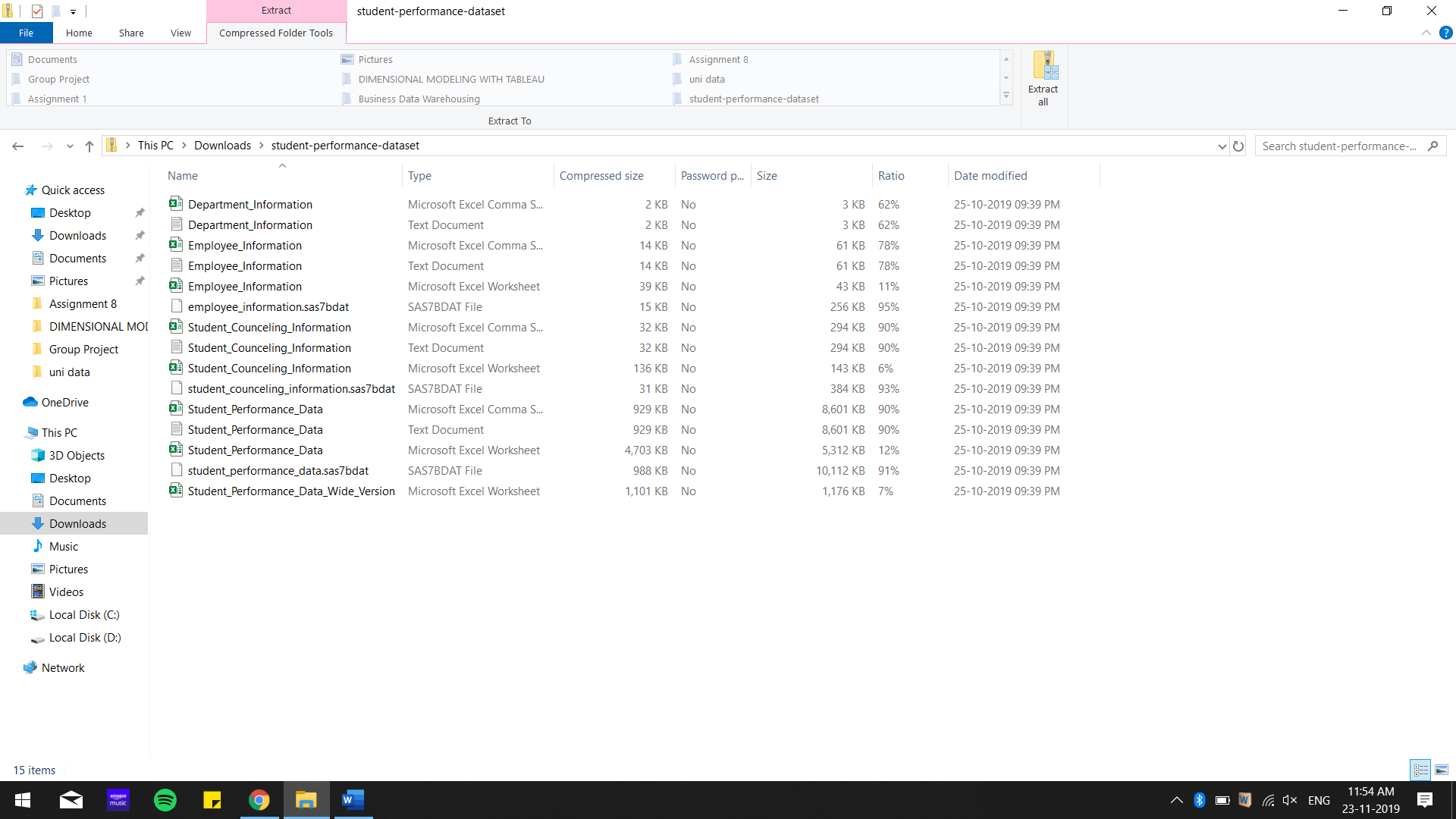
**Dimensional Modelling Using Tableau**

Step by step process of dimensional Modelling in Tableau:

1. **Data Sourcing:**

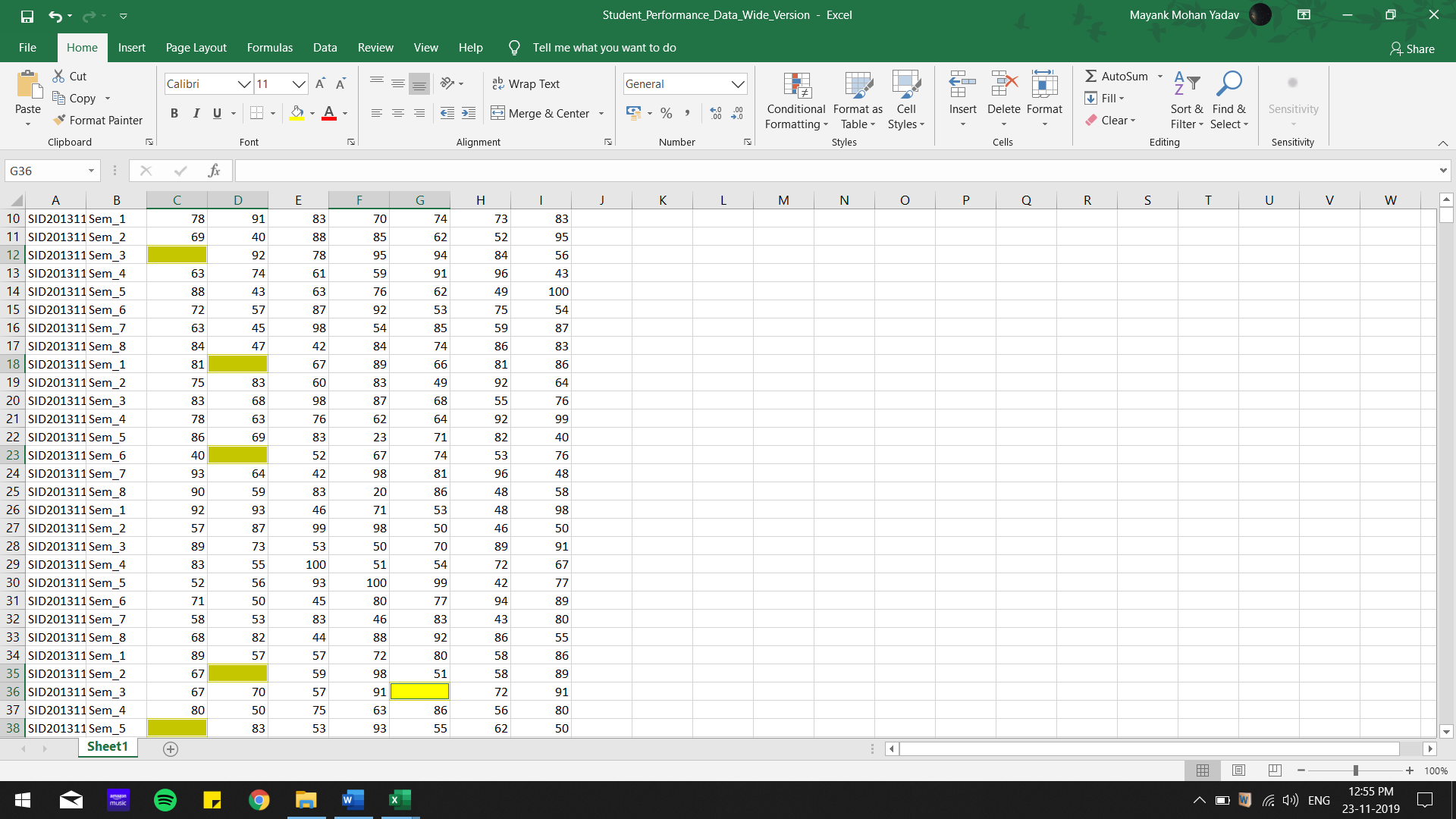
<https://www.kaggle.com/ananta/student-performance-dataset>

We have sourced our dataset from kaggle.com

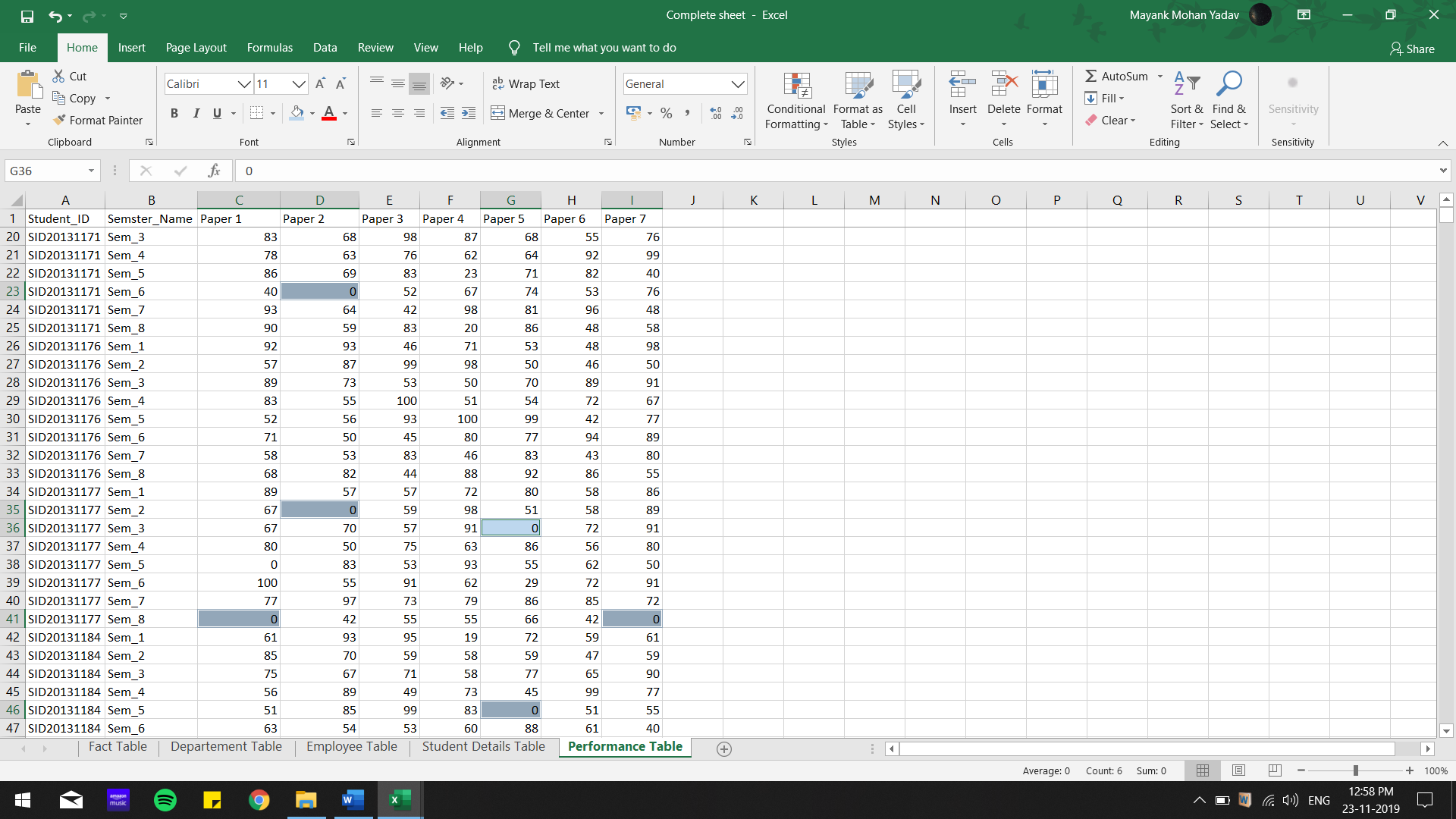


1. **Data cleaning**

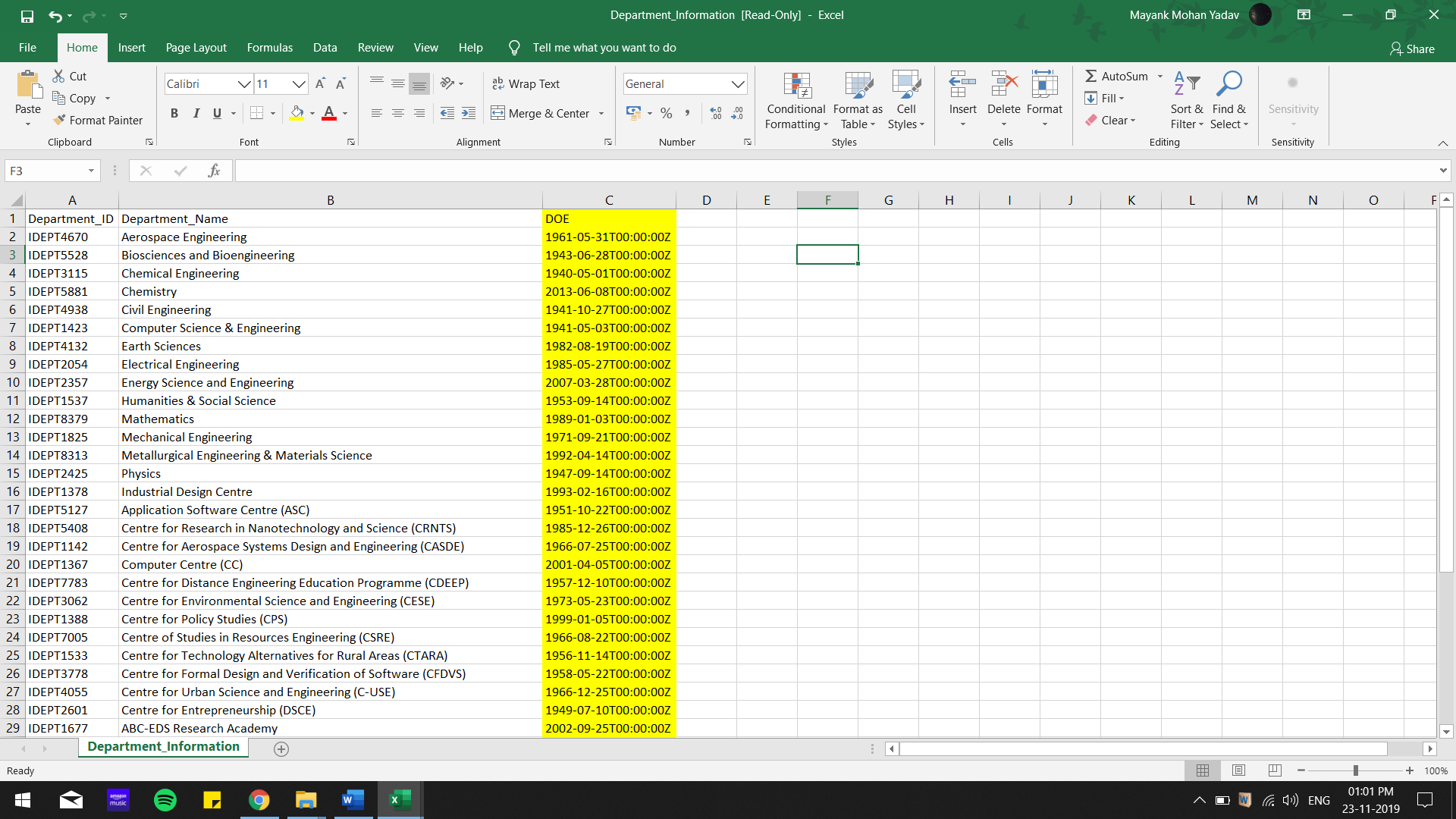
Areas we have touched upon are, dealing with null entries, and removal of unused dimensions.

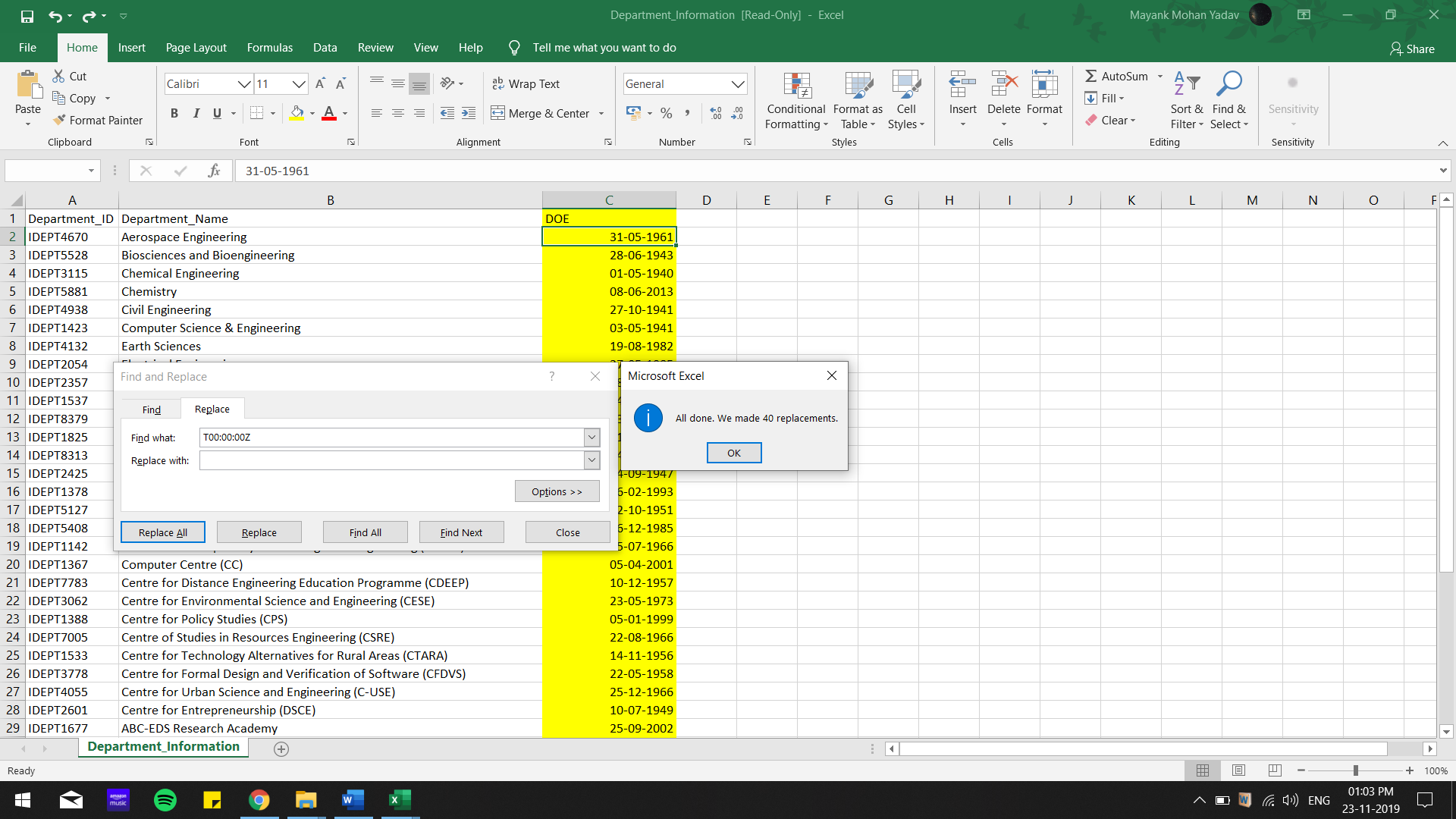


For the purpose of our visualization we have replaced null values with 0.



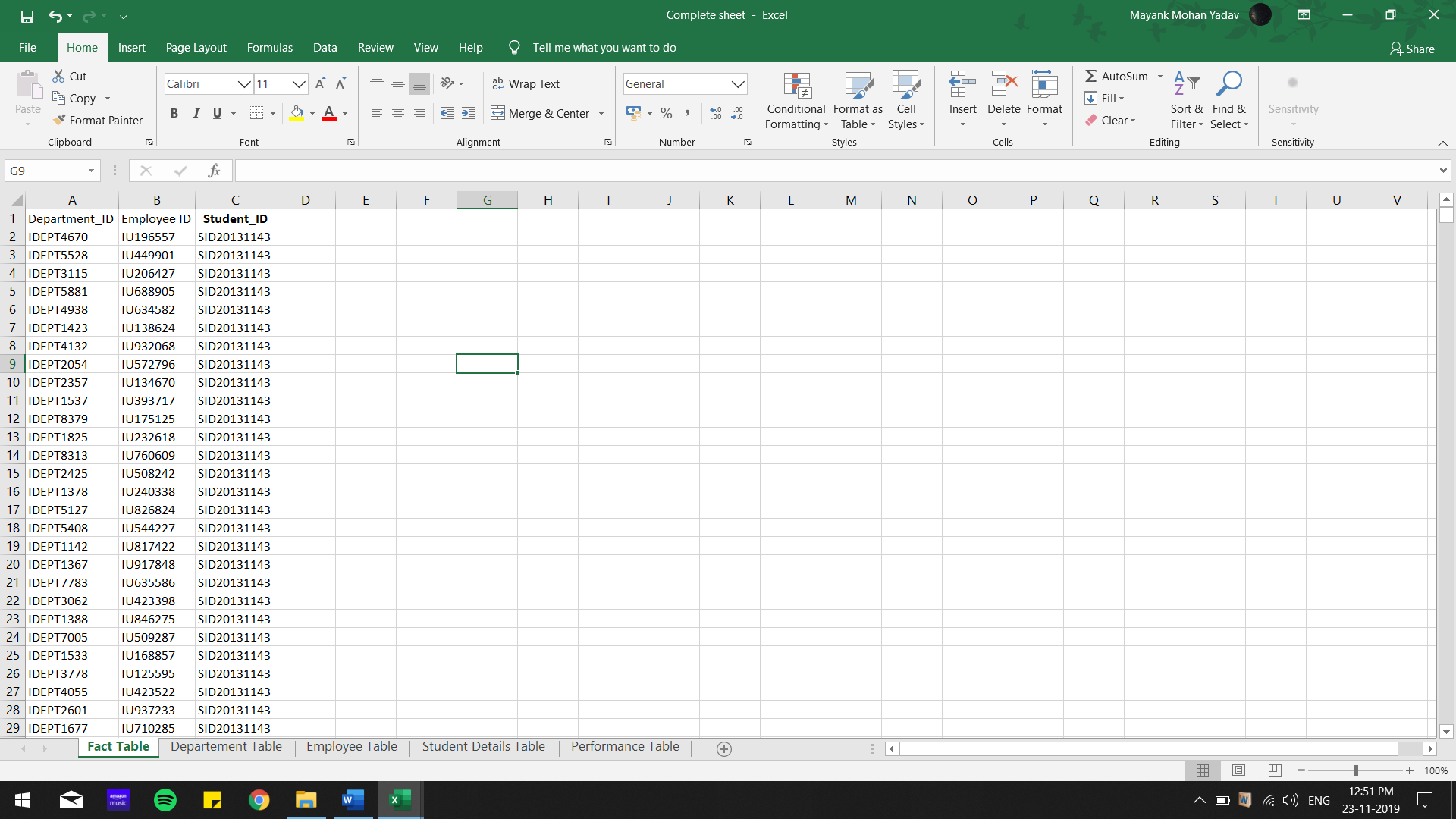
To make the structure of our dataset simpler we have augmented the columns by removing useless information.





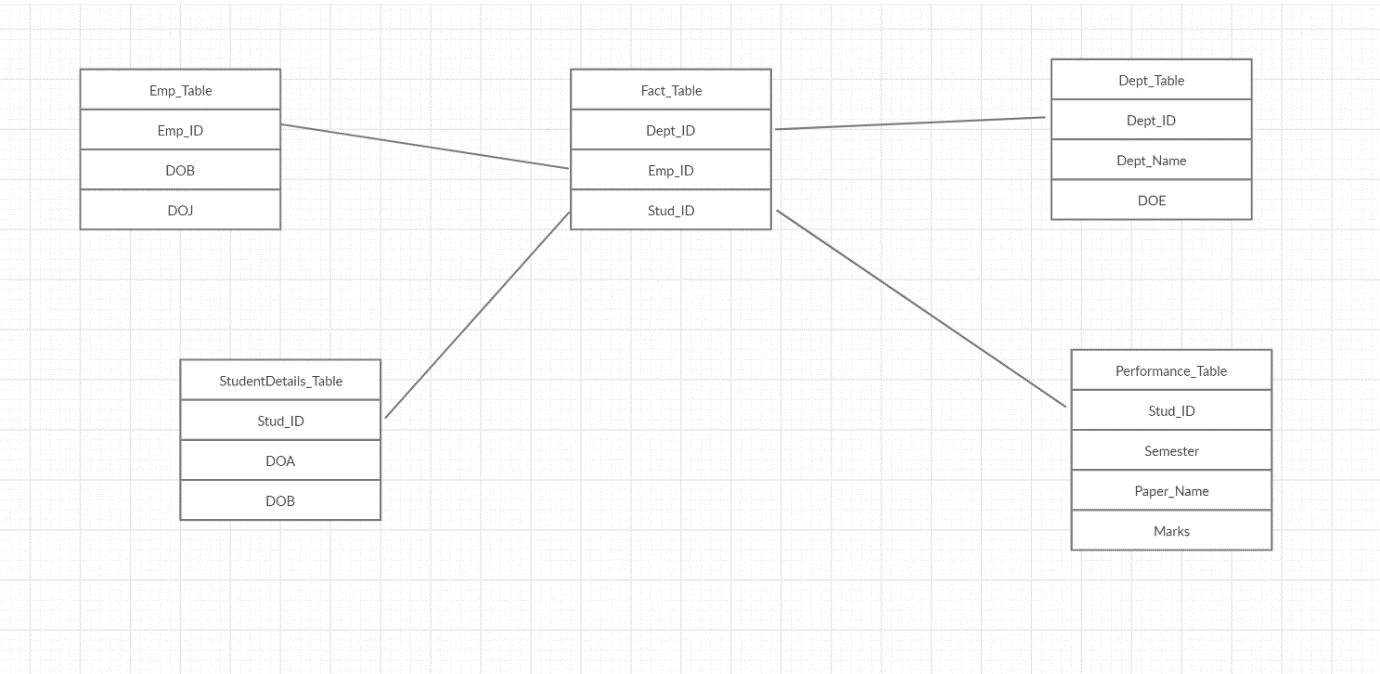
All cleaning is done through MS Excel.

We have consolidated our data files into one excel sheet.



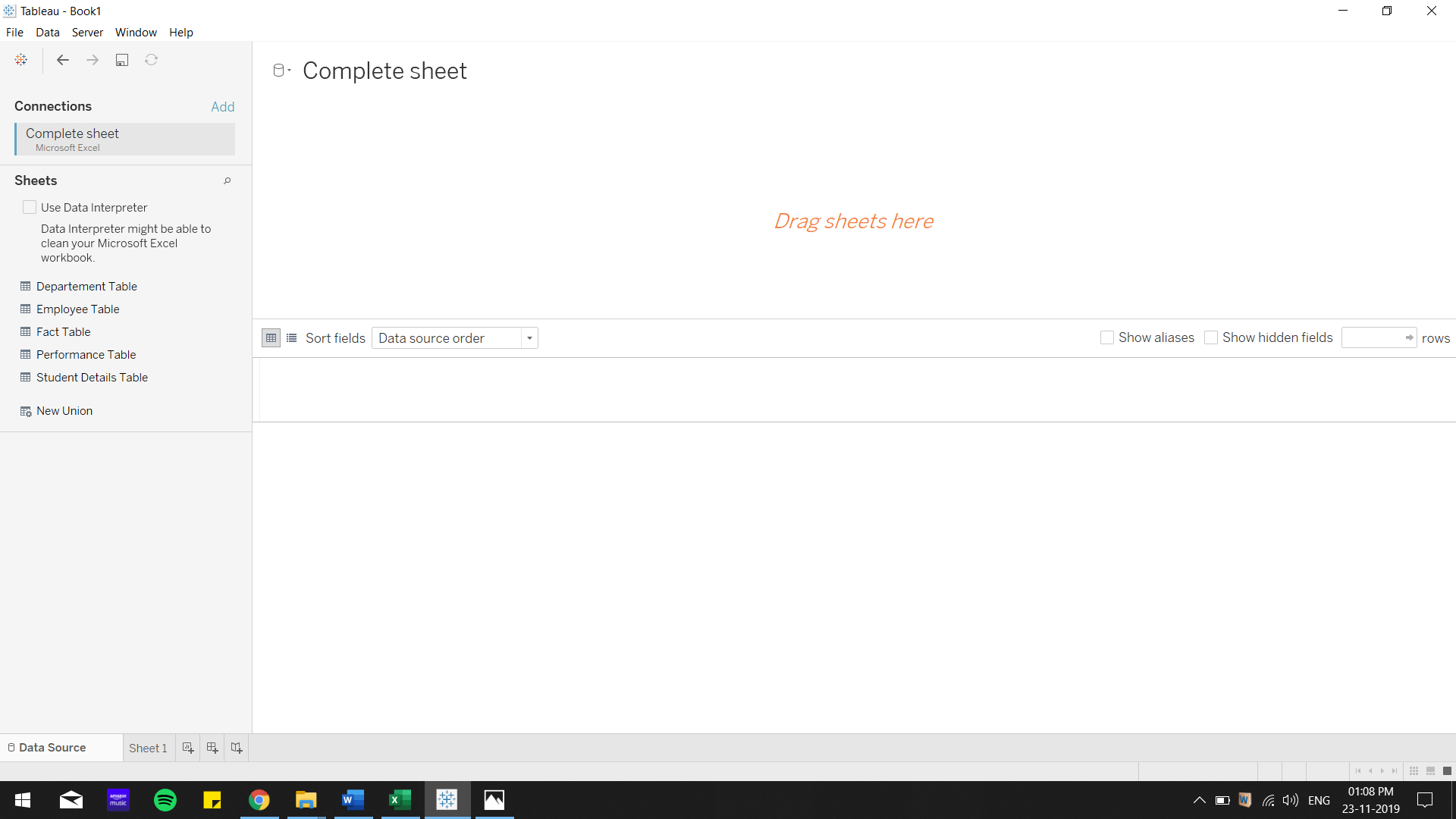
1. **Data Modelling in Tableau**

We have created a model for our dataset with the help of an Entity-Relationship diagram.

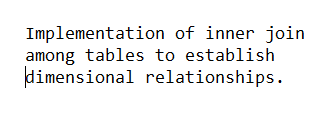
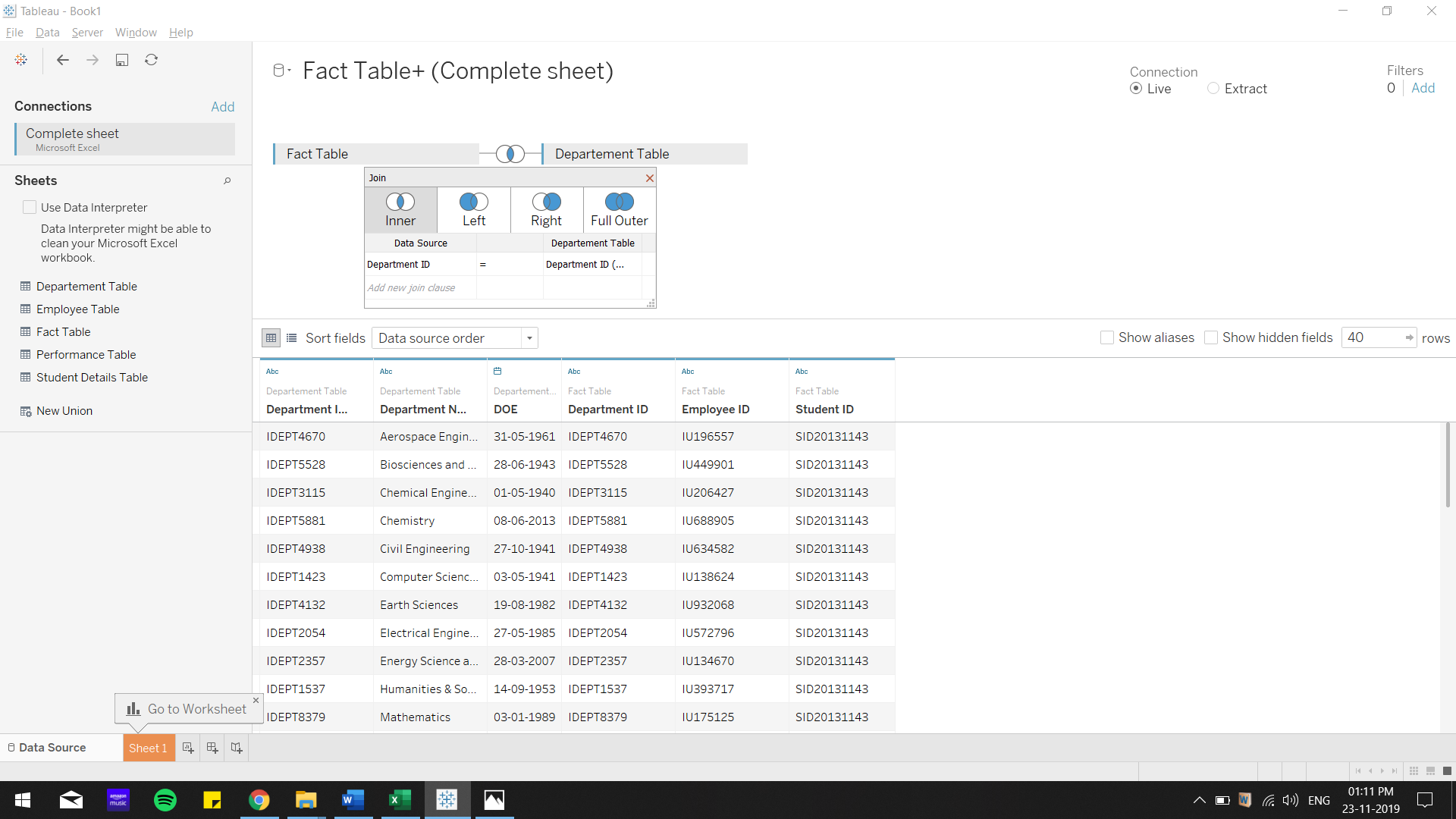


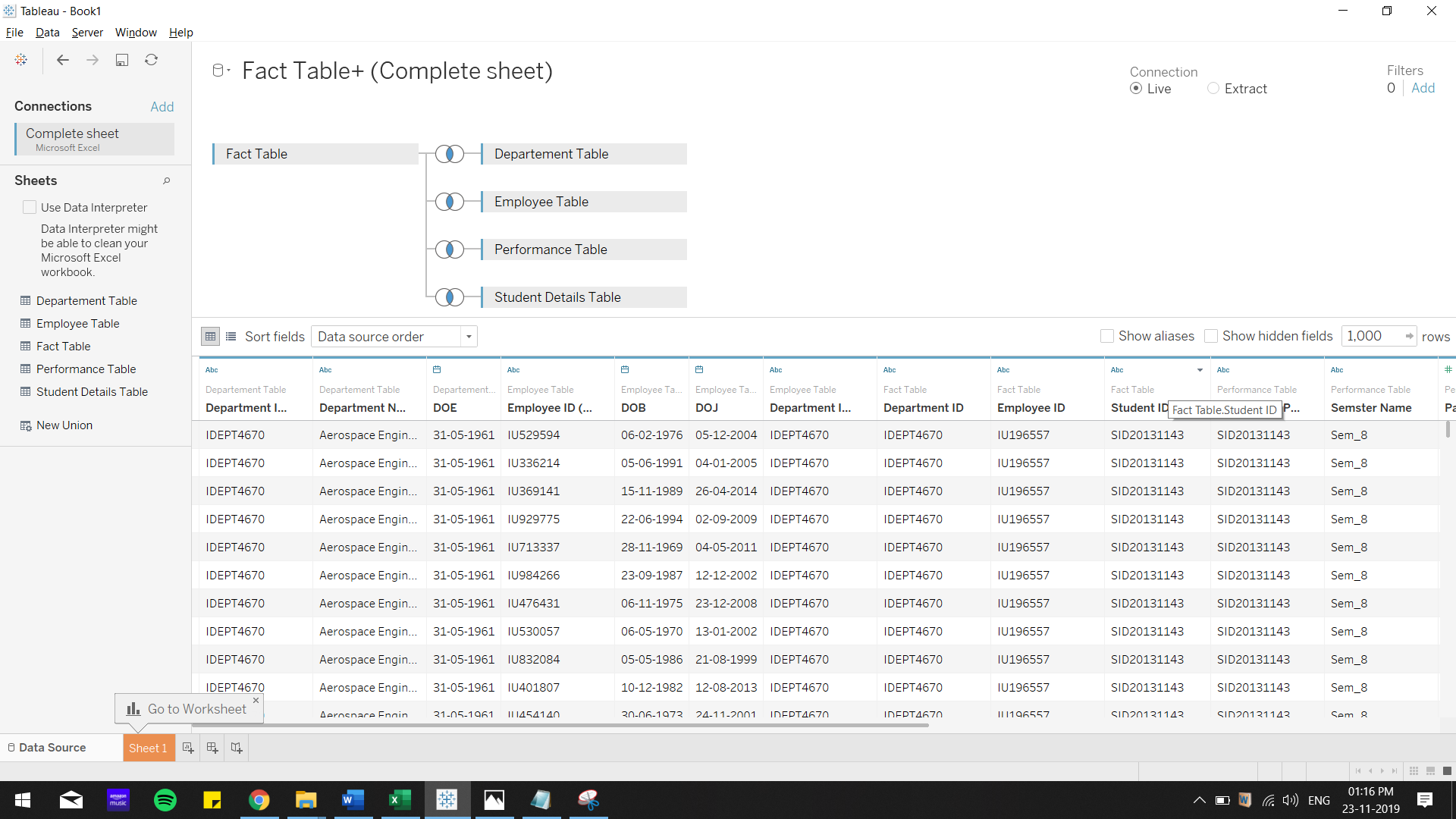
Now we proceed to visualize this model in Tableau.

We have imported our Consolidated sheet in Tableau

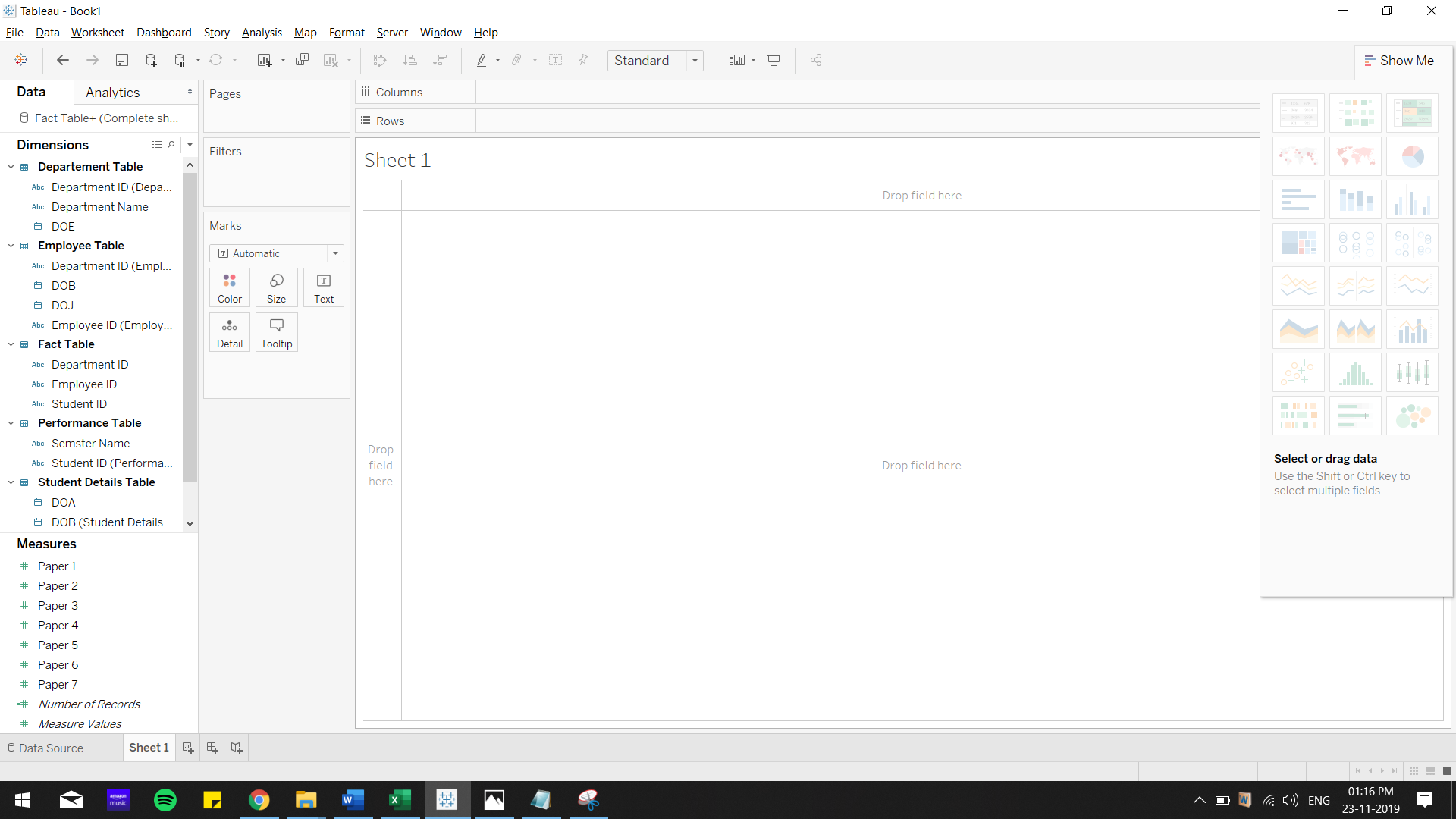


To establish relationships among dimensions we join our tables using inner join.

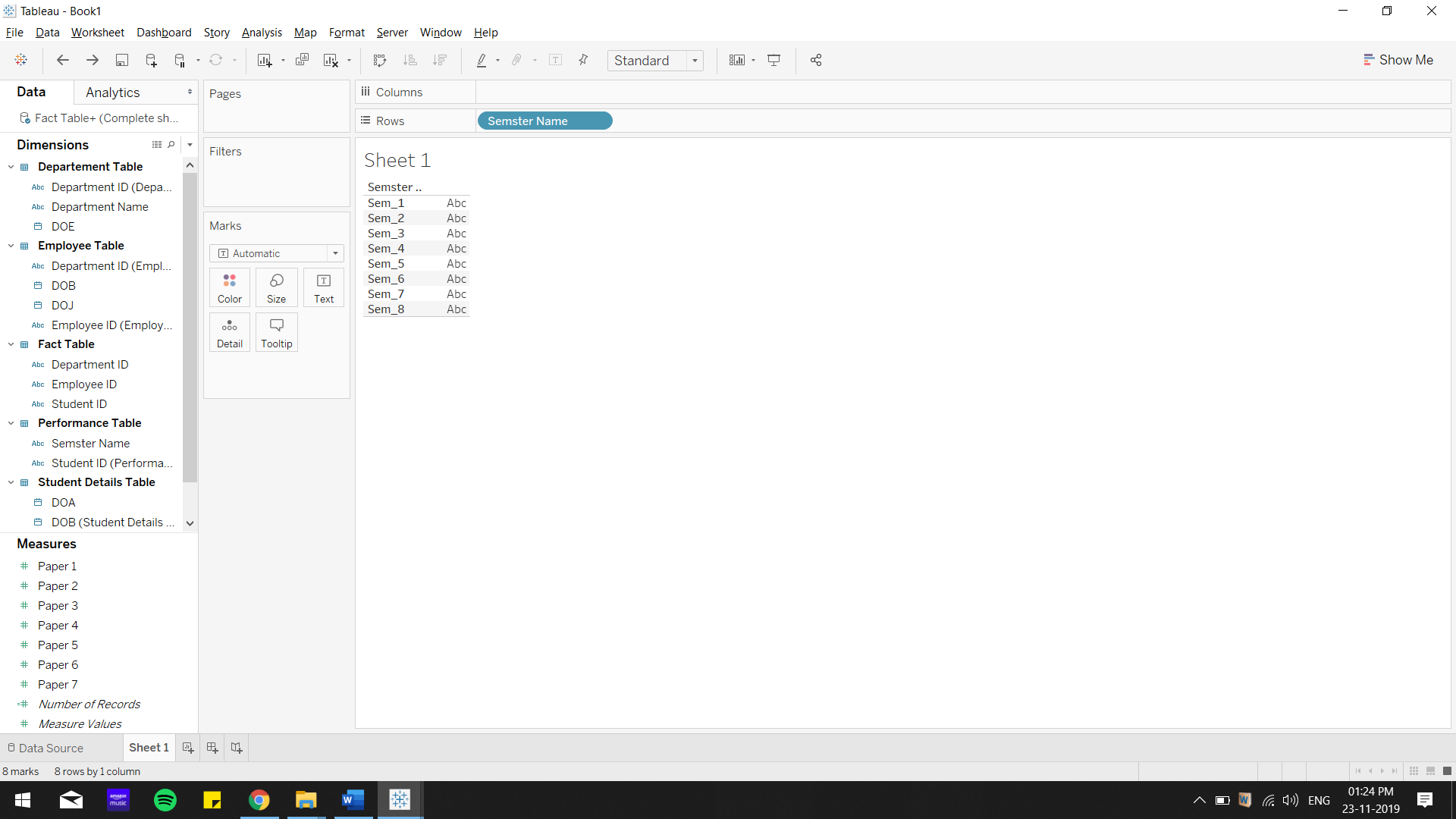


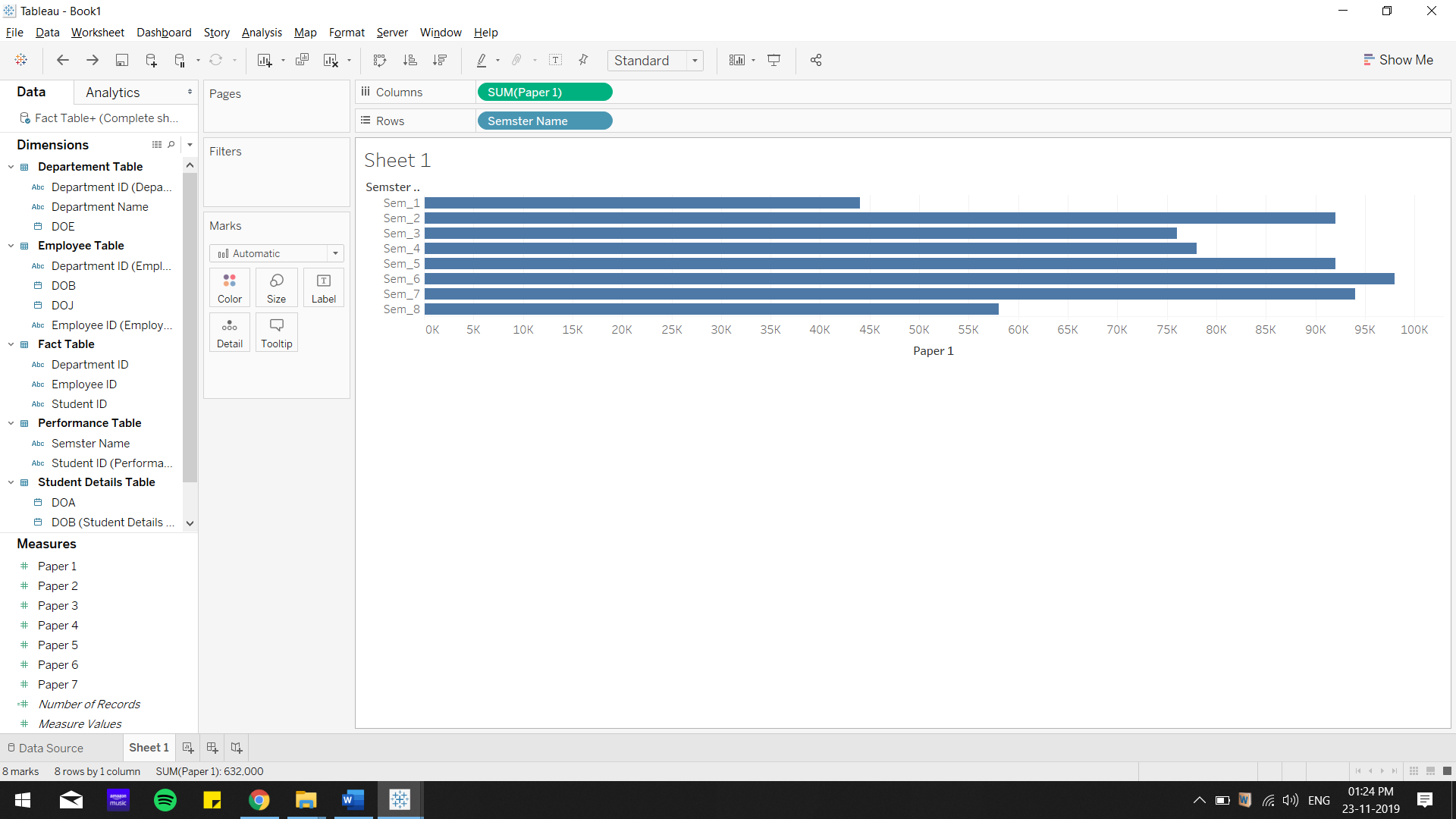
**Visualization part:**

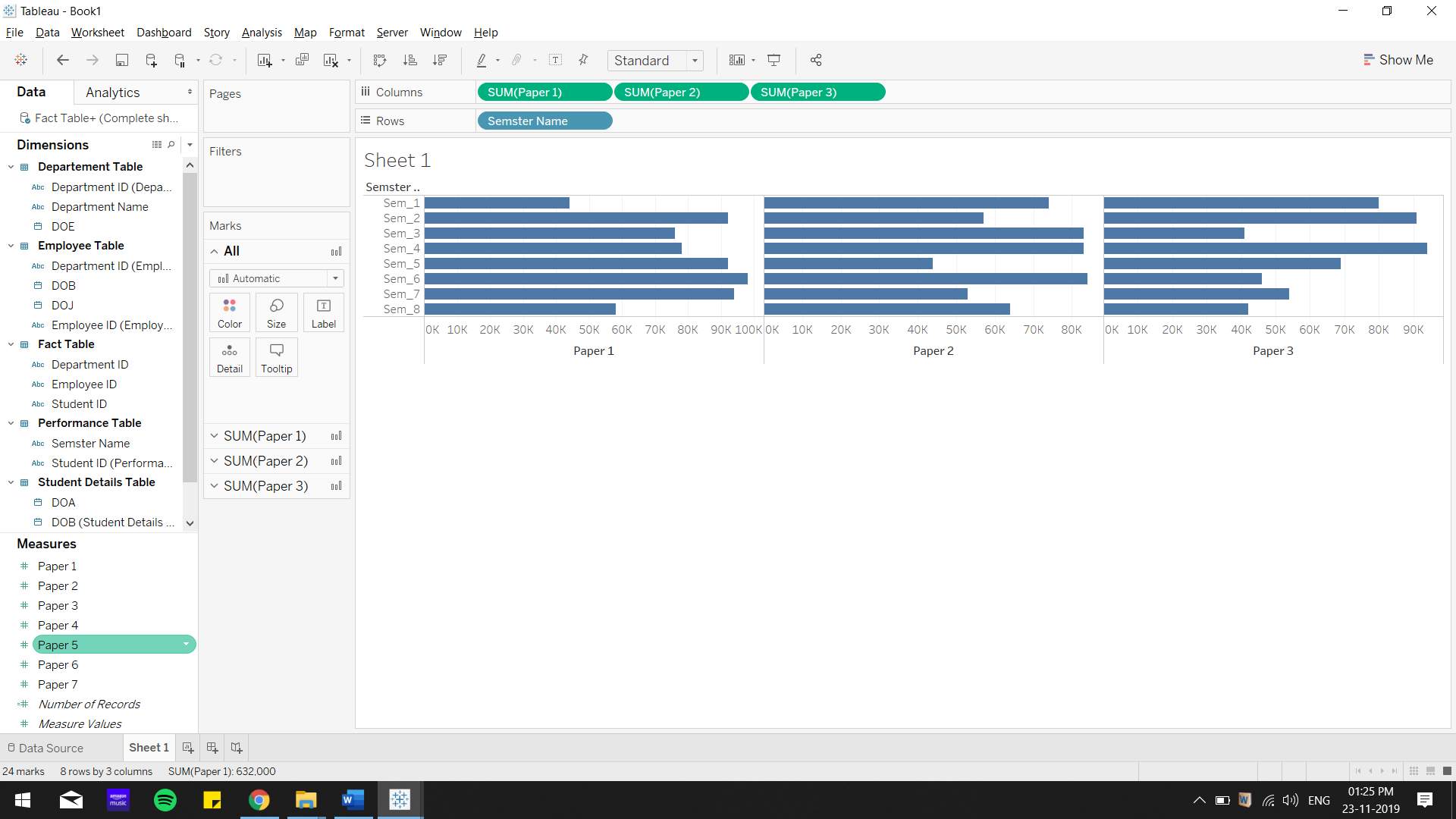


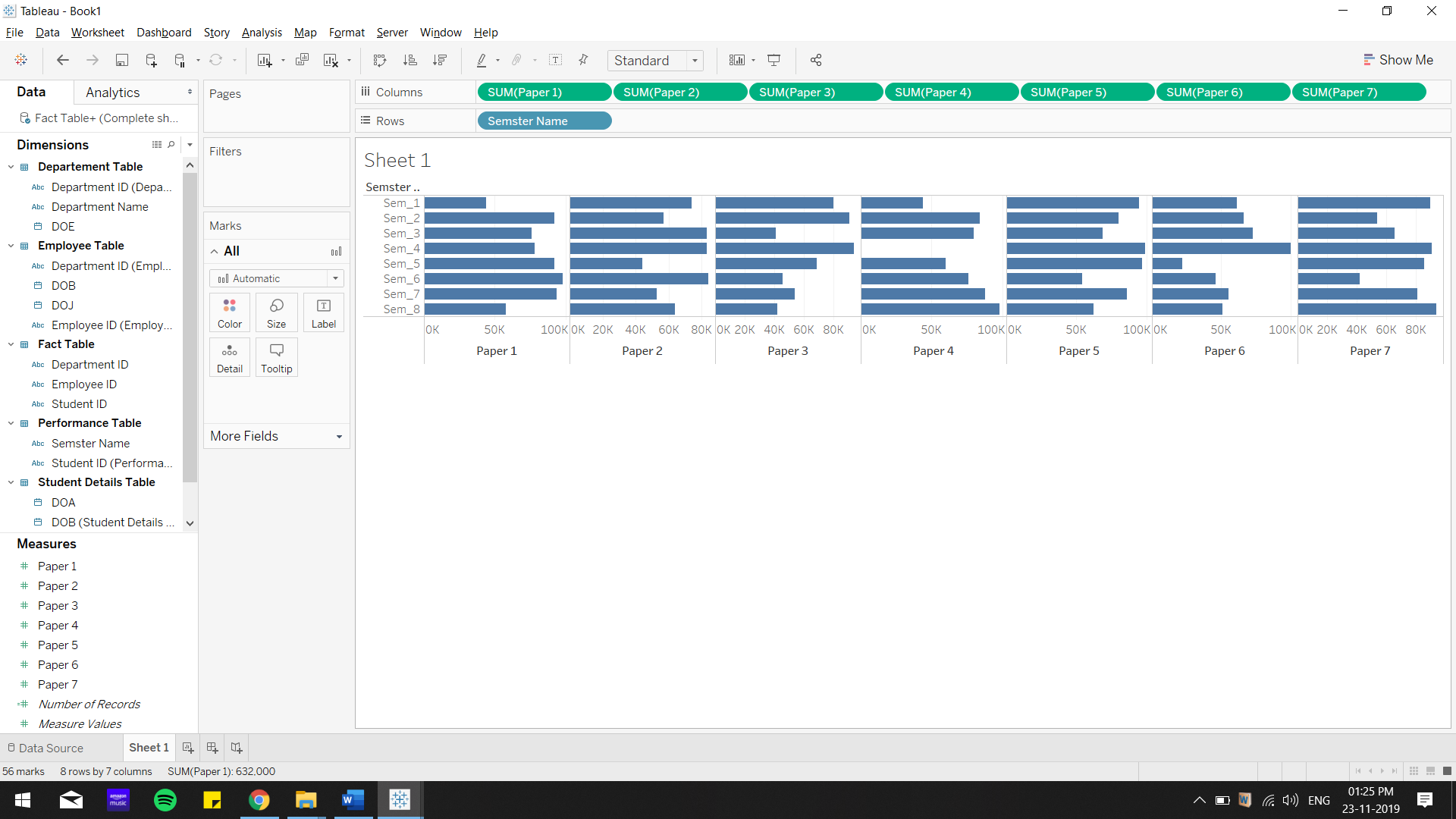
**Insight 1:**

Determining which semester was easiest and hardest to score:

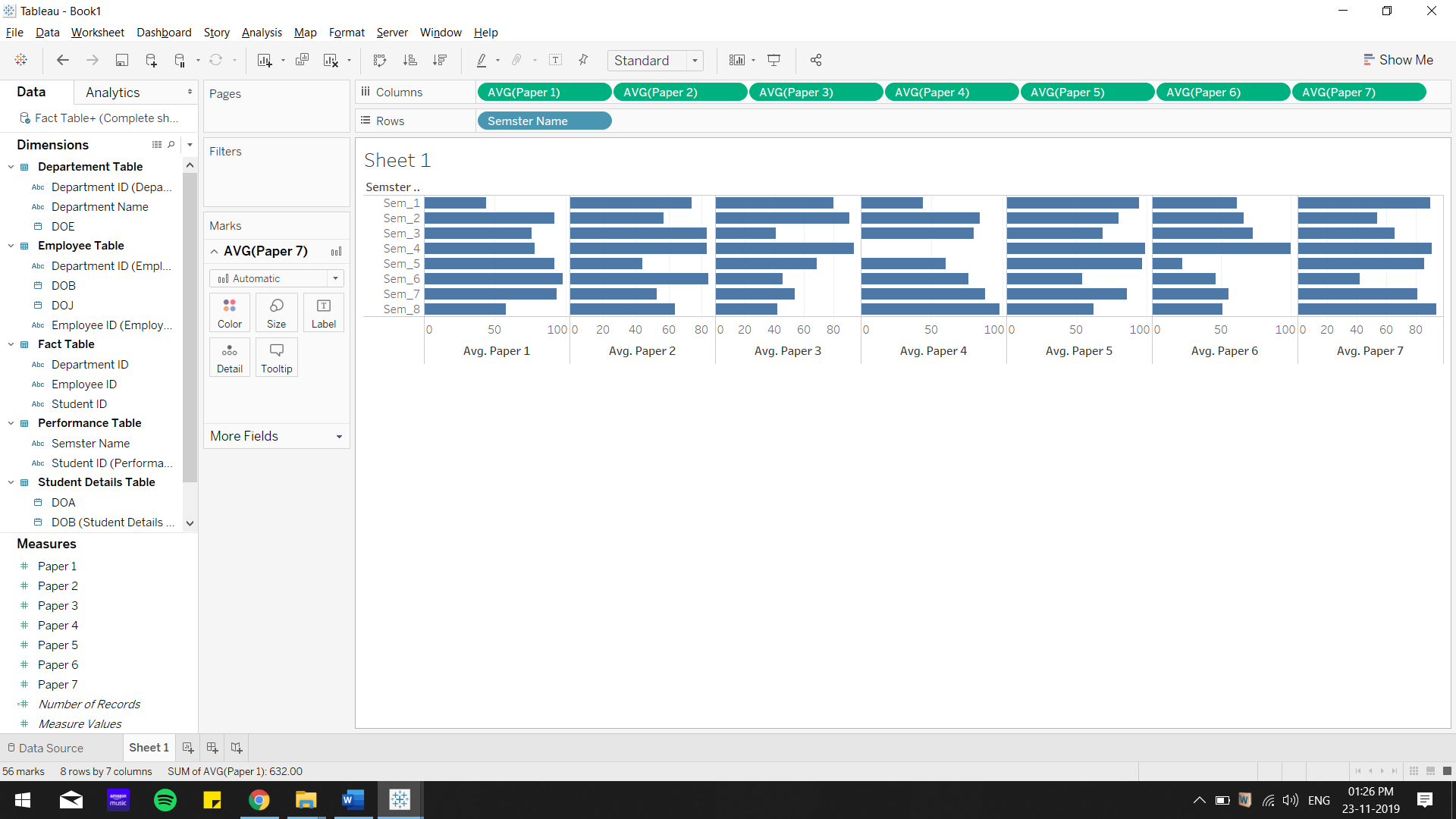




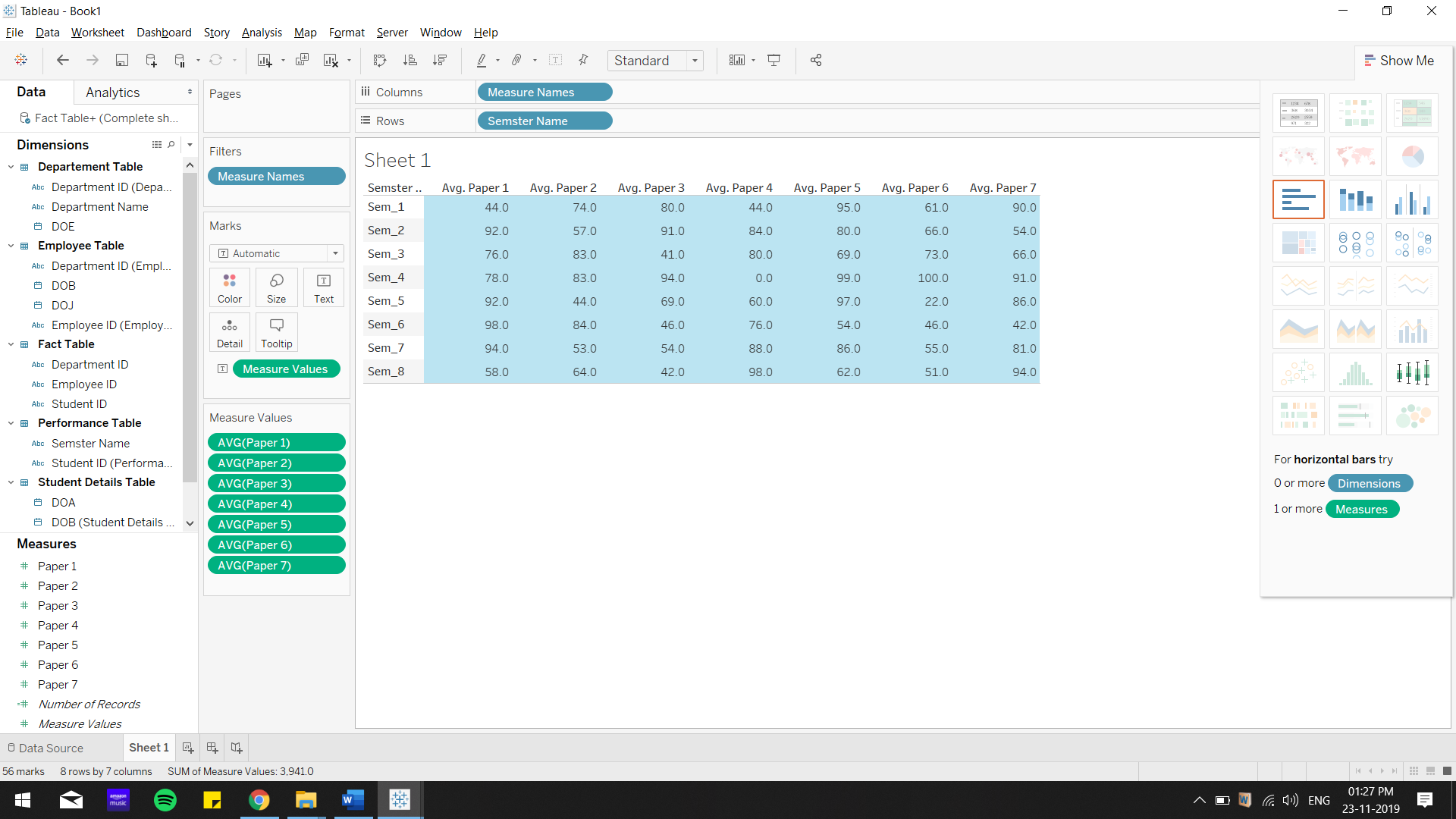




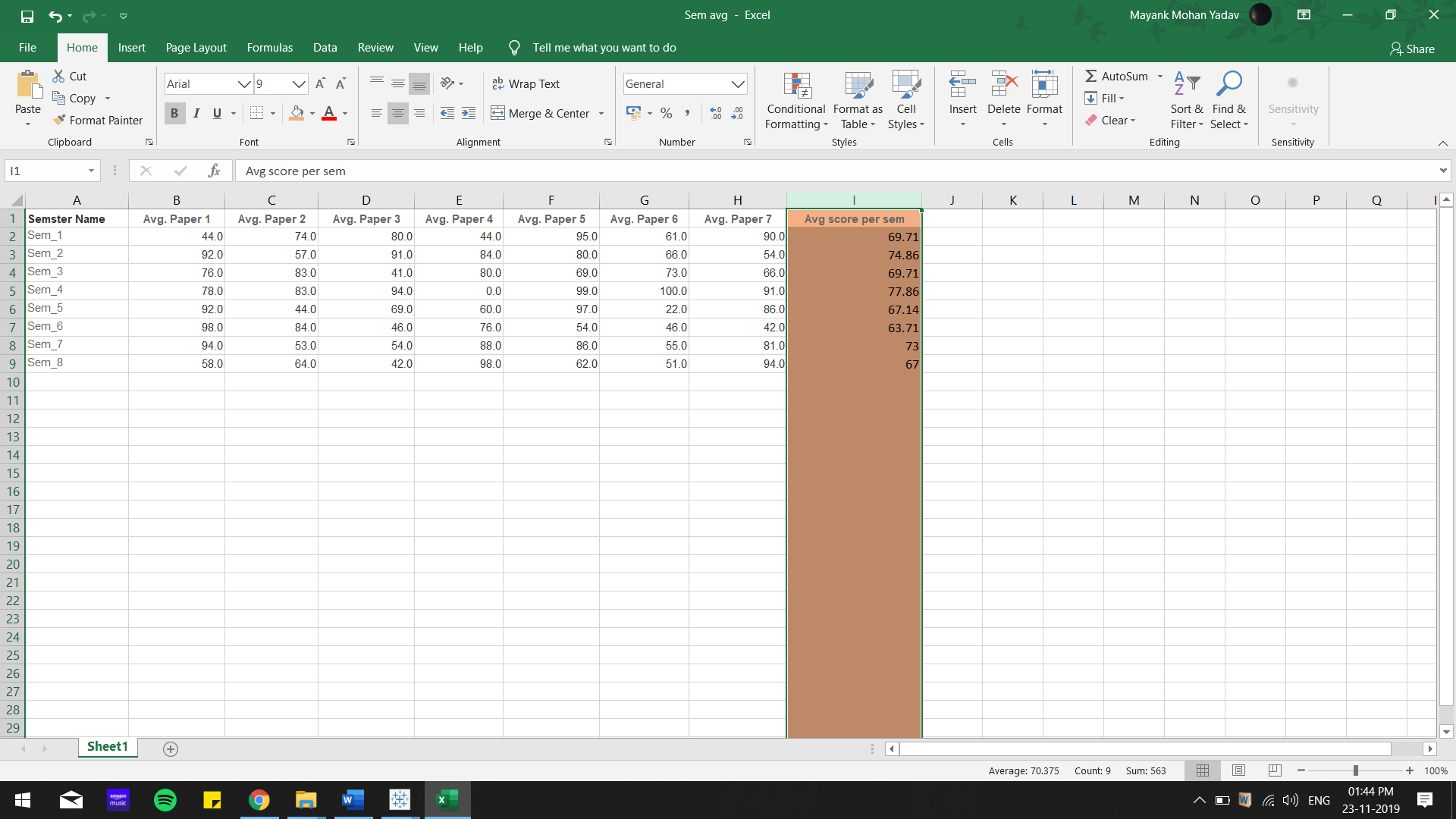
Now changing the aggregation from sum to average:

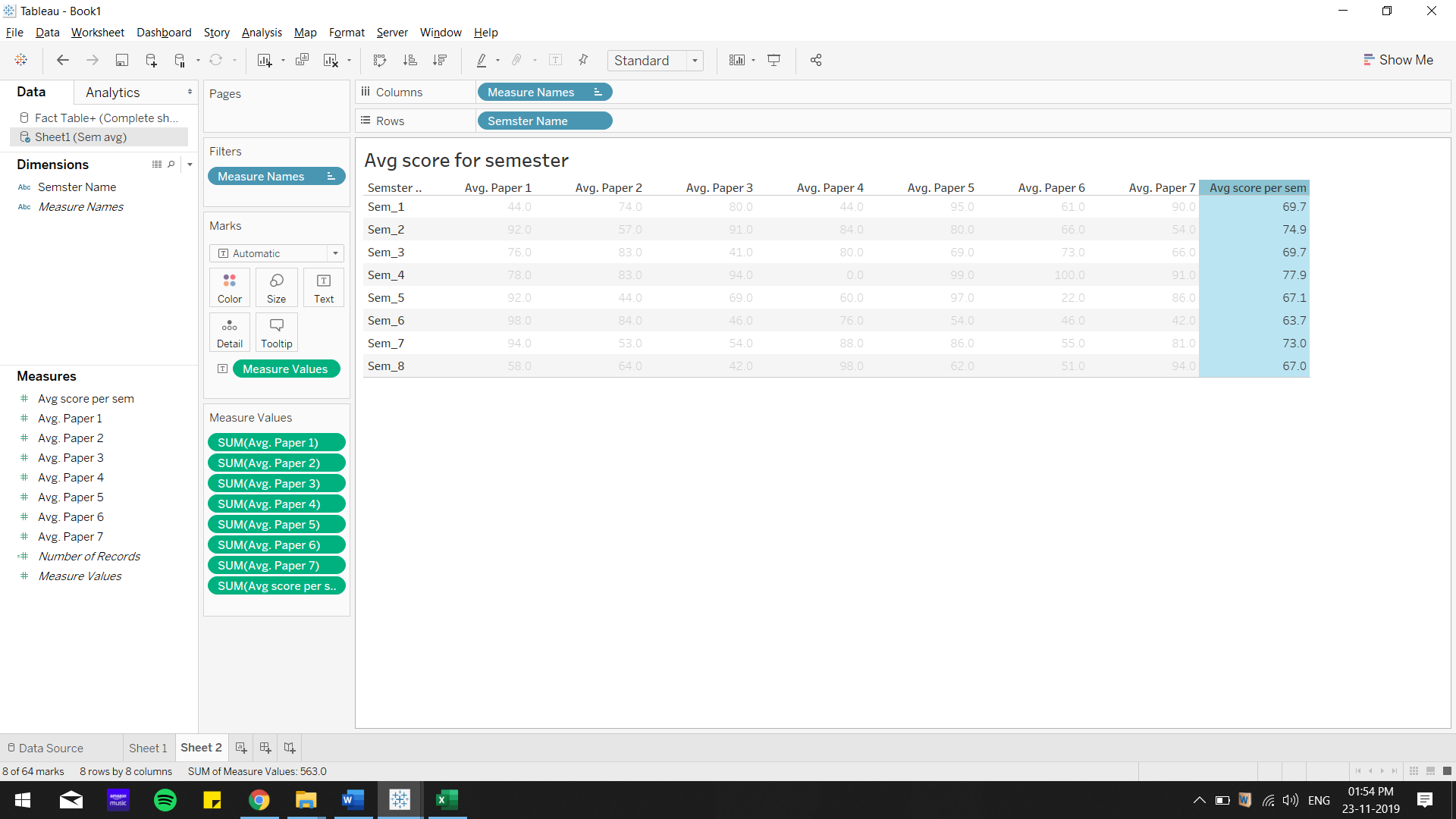


Now changing horizontal bar chart to text table:

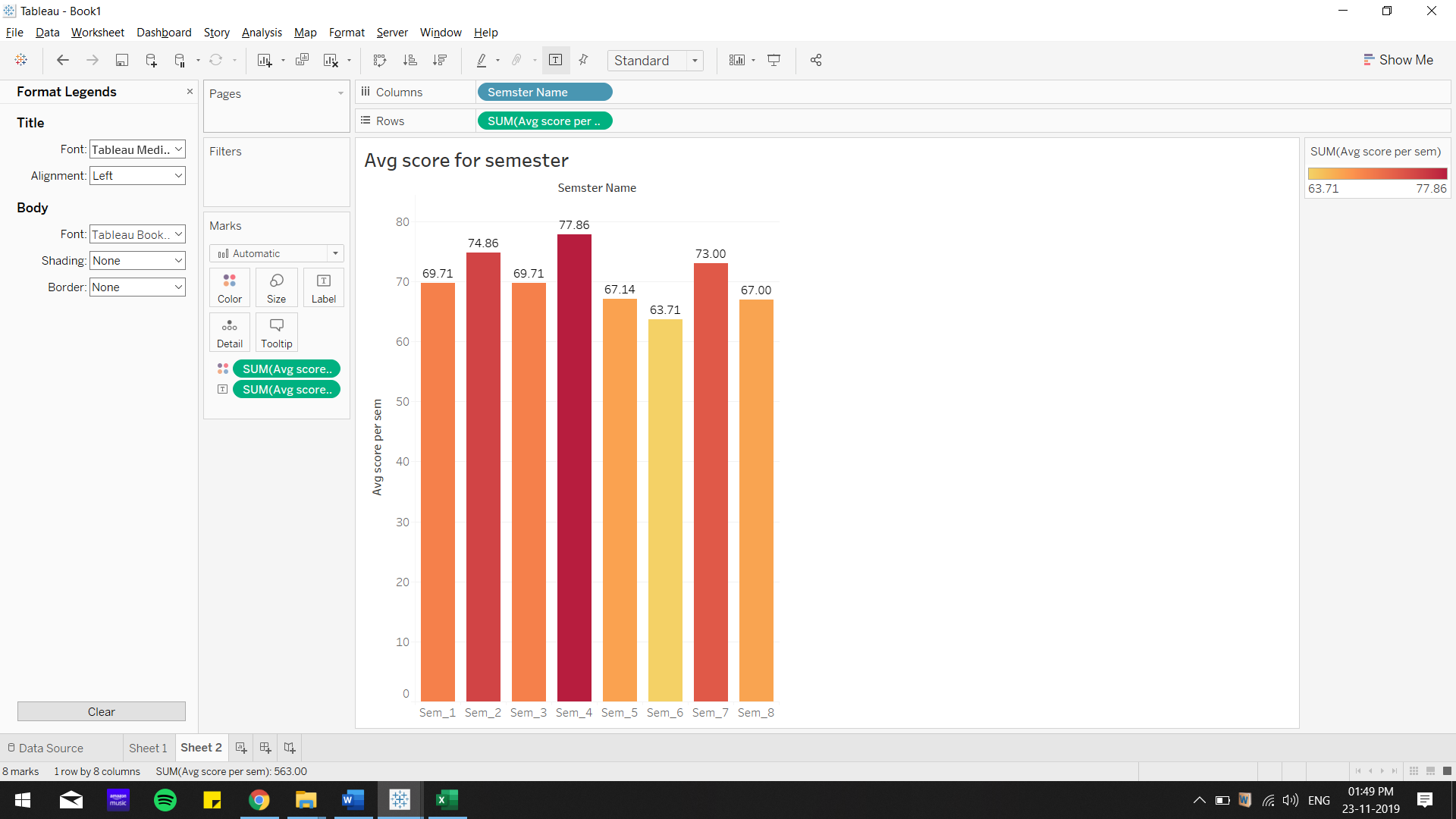


Now calculating the average performance for each semester:





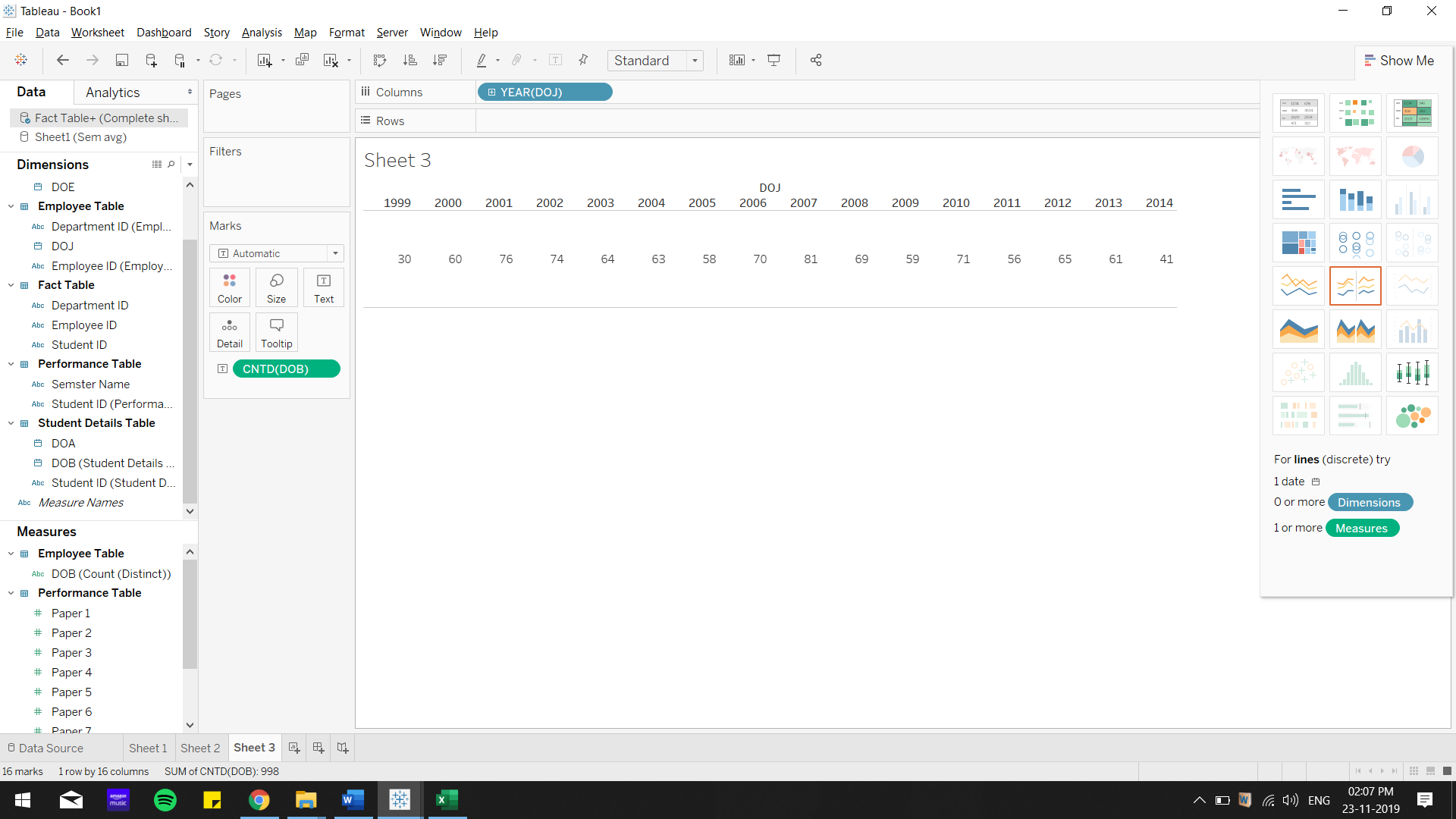
Visualizing the above result in Tableau:



From above results we can conclude that:

1. **Semester 4 was easiest to score with an average of 77.86 marks in each paper**
2. **Semester 6 was hardest to score with an average of 63.71 marks in each paper**

**Insight 2:**Comparing number of employee hires to the university with respect to years.



Visualizing it through a Horizontal bar chart:

Convert DOJ to string in order to change it to a measure for our visualization. Drag it to Rows. Drag DOB to columns to get number of people who have joined the university from 1999 through 2014.



**Conclusion:**

**From above insight, we can conclude that 2007 was the year with highest number of employee hiring’s (81)**

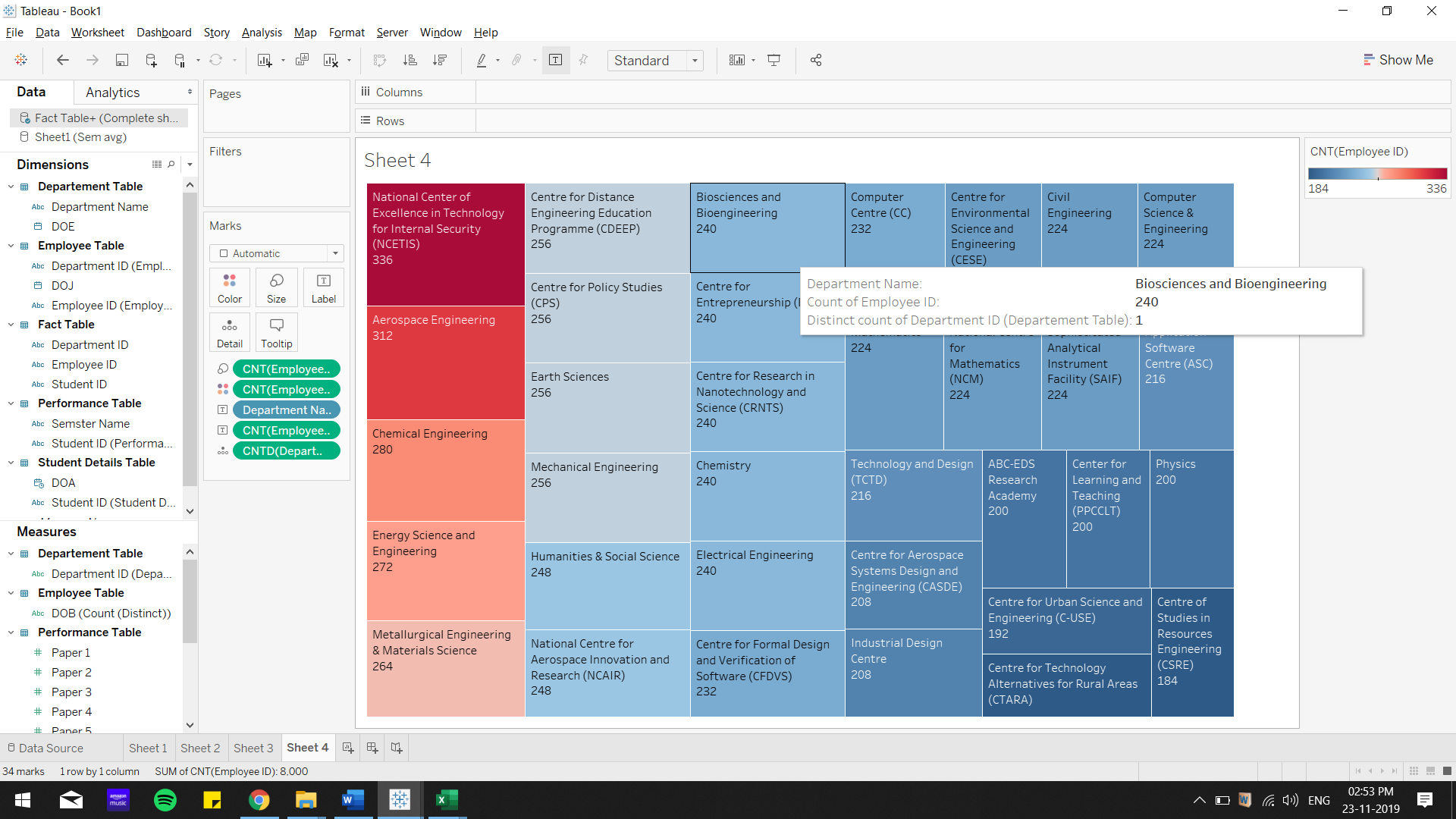
**From above insight, we can conclude that 1999 was the year with lowest number of employee hiring’s (30)**

**Insight 3:** Distribution of employees across the departments.

We drag from Employee\_Table, Department name to columns, and Employee\_ID to rows(distinct count)

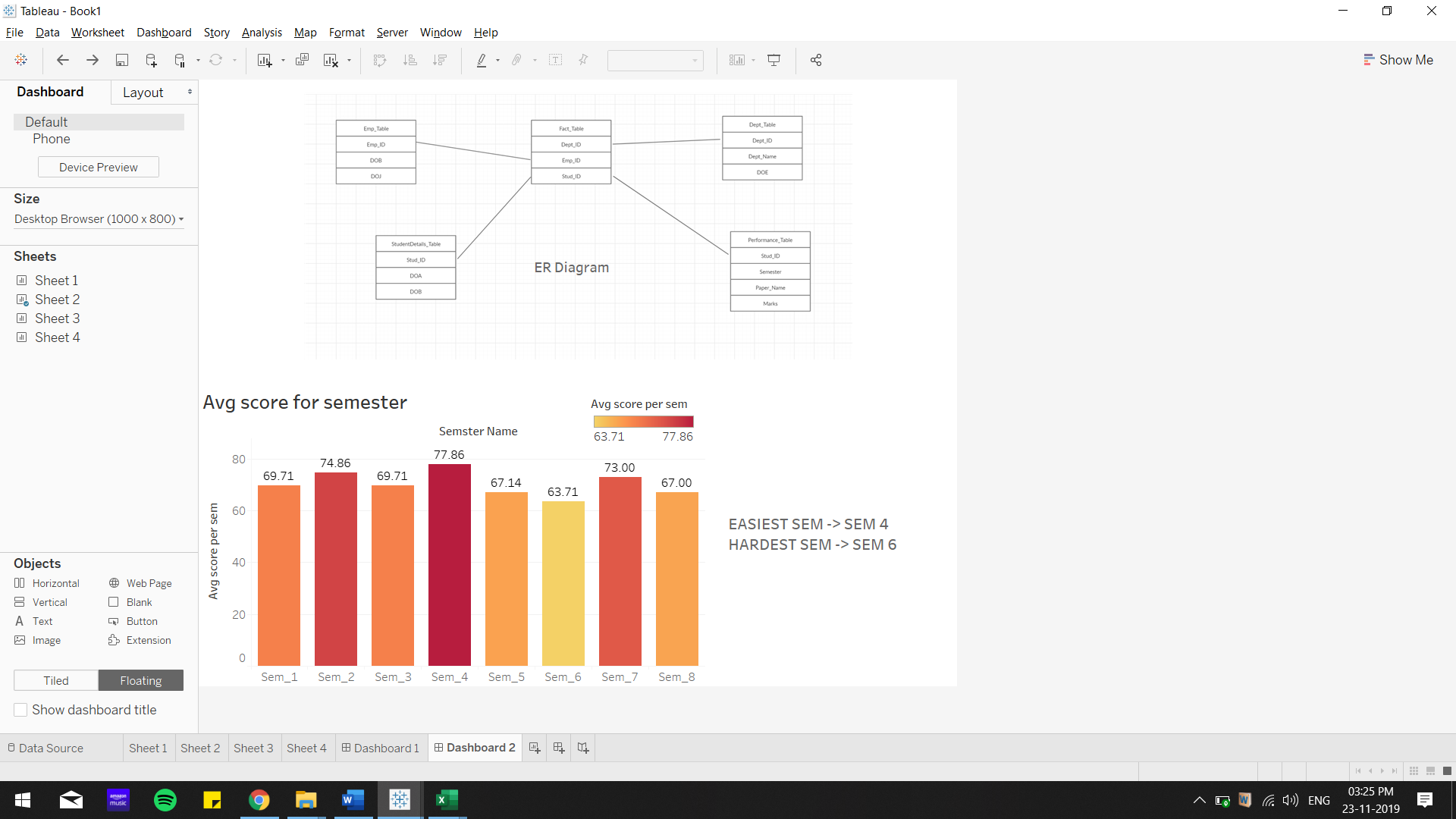
For this, we use 2 dimensions: Department name, and distinct count of(Employee\_ID).

For a coherent visualization we use Treemaps with colour gradient ranging from blue(lowest number of employees) to red(highest number of employees) across departments.



**DASHBOARD VIEW:**

**DASHBOARD 1: In this dashboard, we have an ER diagram of the data model and based on it we have analysed the difficulty of each semester.**



**DASHBOARD 2: In this dashboard, we have horizontal bar chart showing the number of hires with respect to years. The tree map here gives the number of employees working for particular department.**

