**Course: Data Mining**

**Homework 1**

**Student ID: 20C14001**

**Student Name: Le Duong Tuan Anh**

**Dataset: “top\_movies\_by\_title.csv”**

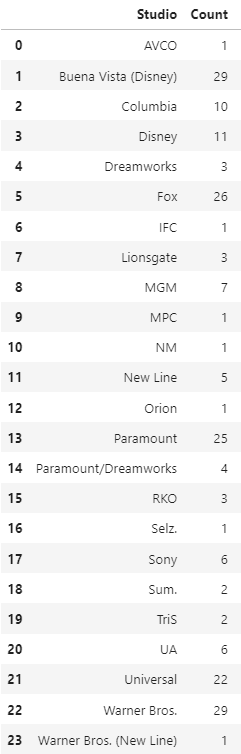
*1) Desbribe the dataset*

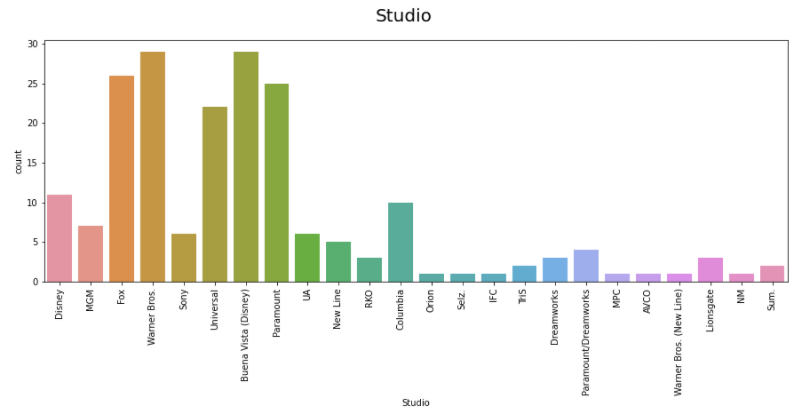
Dataset type: Record.

This dataset has 200 rows and 5 columns.

|  |  |  |  |
| --- | --- | --- | --- |
| Column | Type | Datatype | Has missing value? |
| Title | Metadata | String | No |
| Studio | Nominal | String | No |
| Gross | Interval-scaled | Integer | No |
| Gross (Adjusted) | Interval-scaled | Integer | No |
| Year | Quantitative  (continuous interval) | Integer | No |

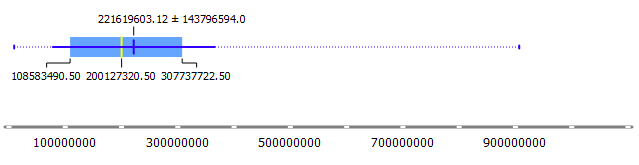
*2) Appy basic statictical descriptions for the dataset*

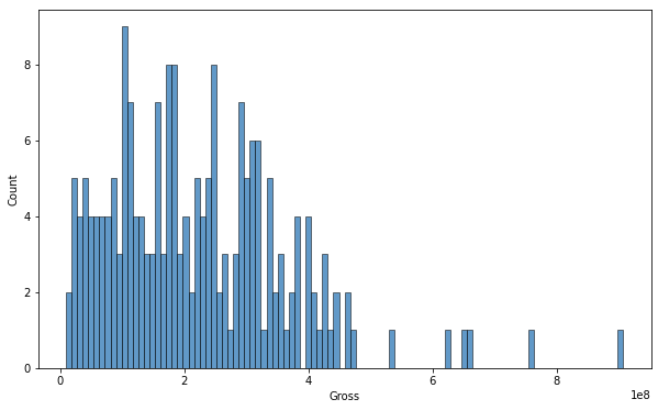
**2.1. Studio**

Distribution Chart

Tag Cloud

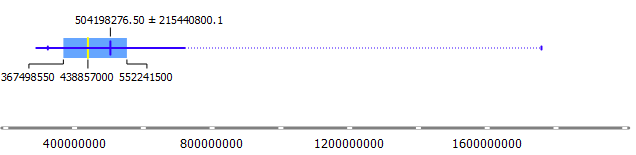
**2.2. Gross**

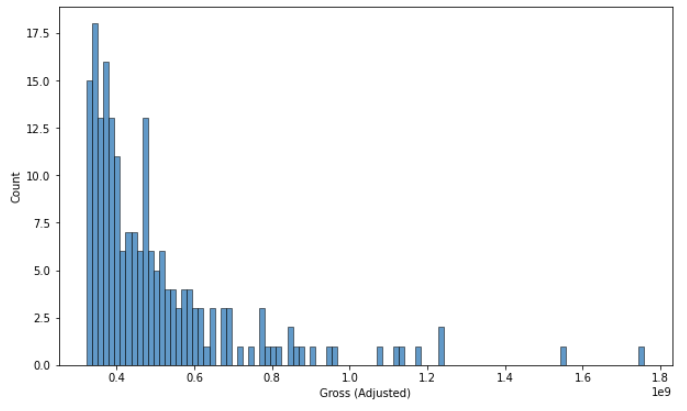
 



**2.3. Gross (Adjusted)**

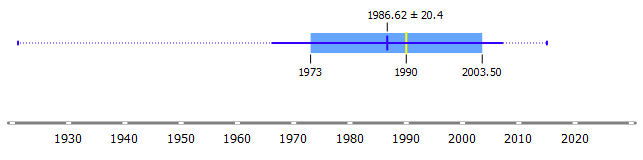


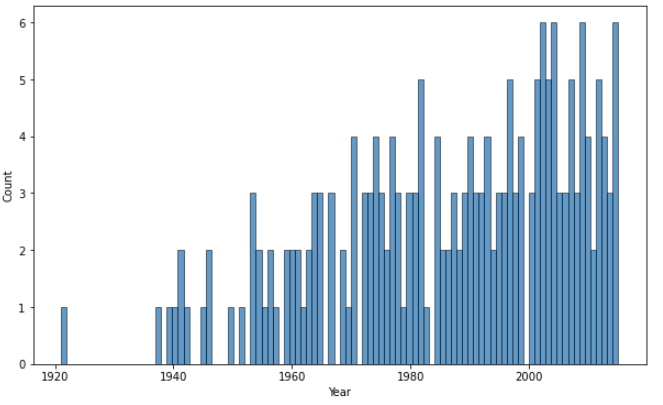


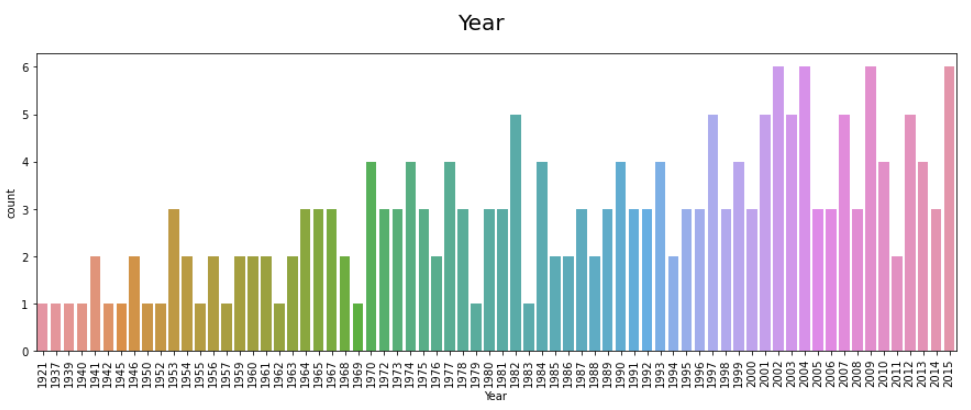


**2.4. Year**

****

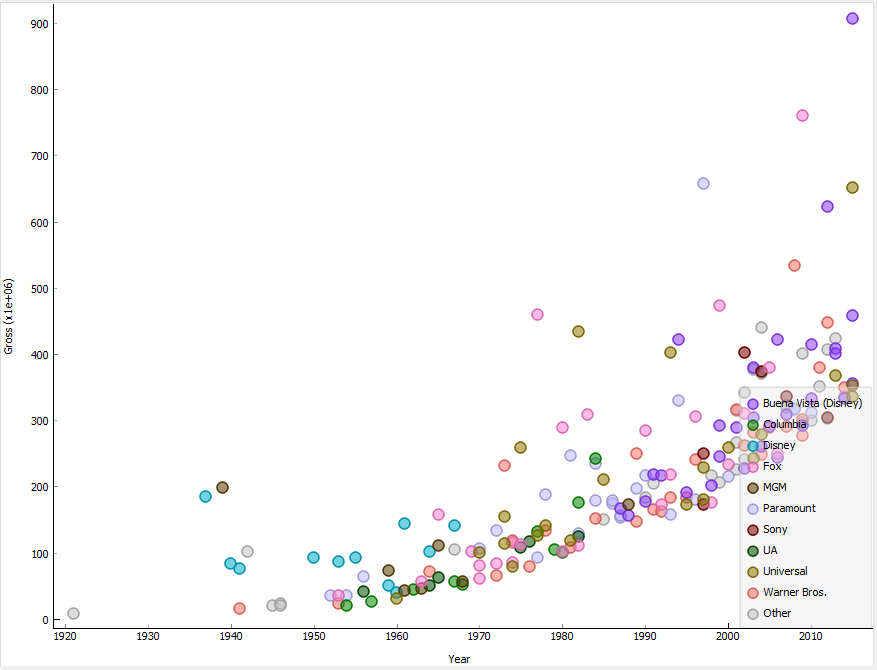
****

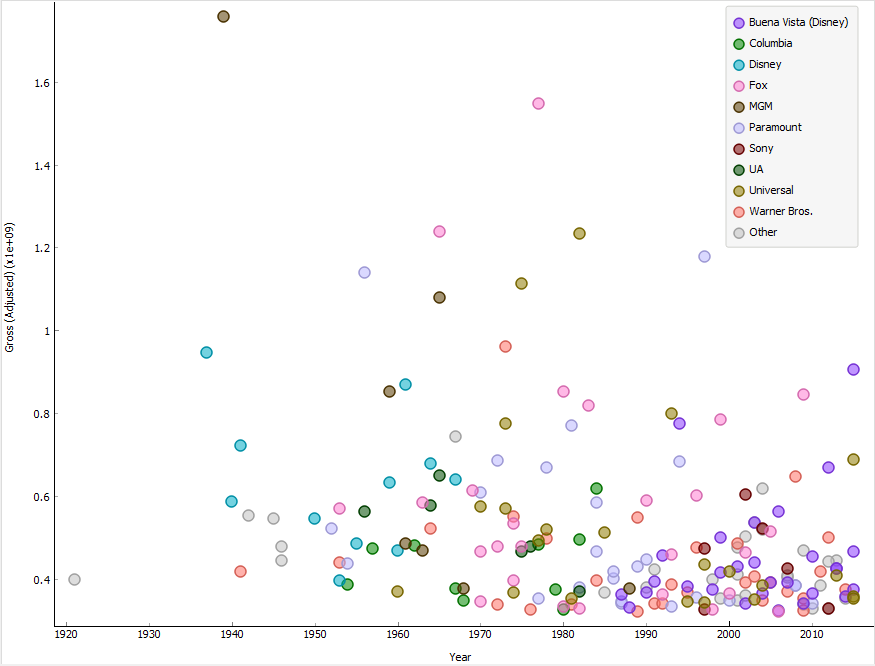
****

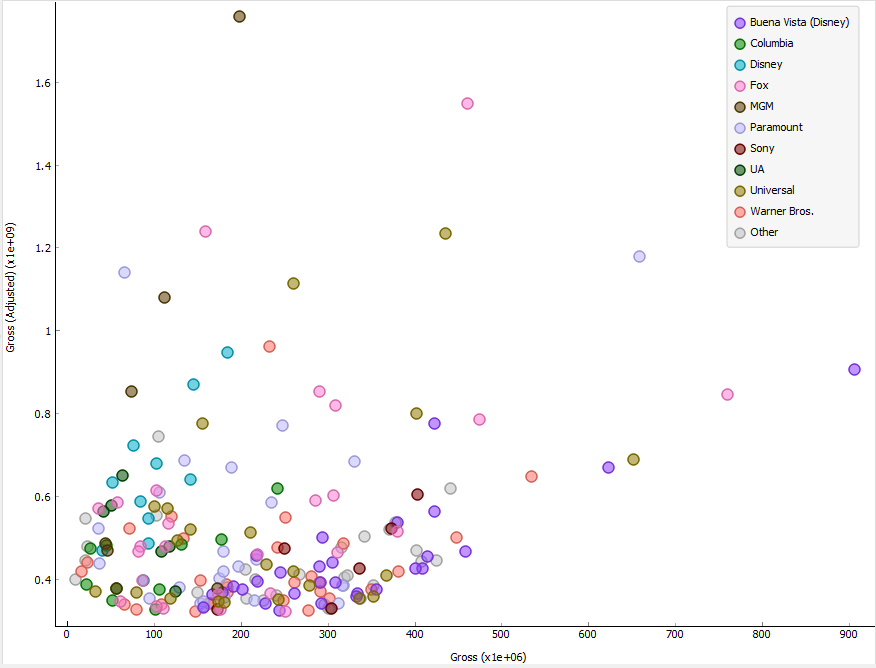
****

*3) Visualize this dataset by using scatterplot matrix*

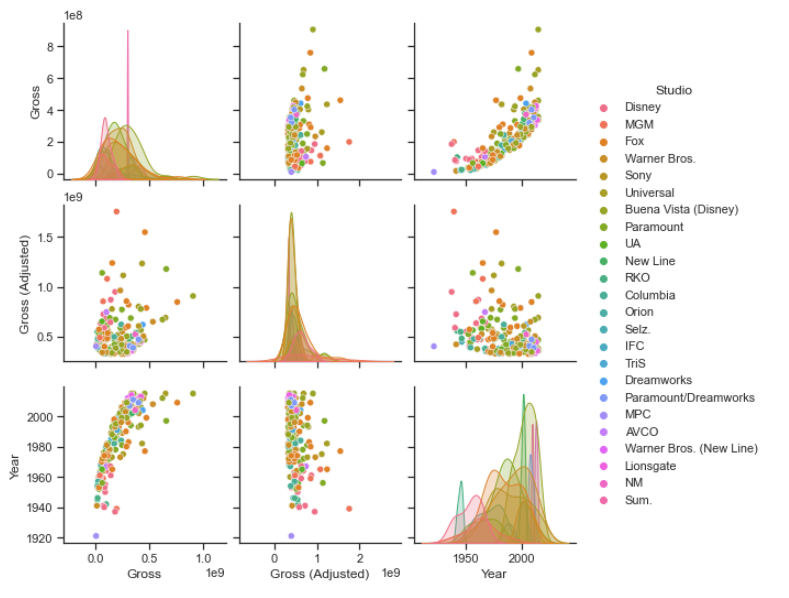
**3.1. For every-pair, distinguish “Studio” column.**



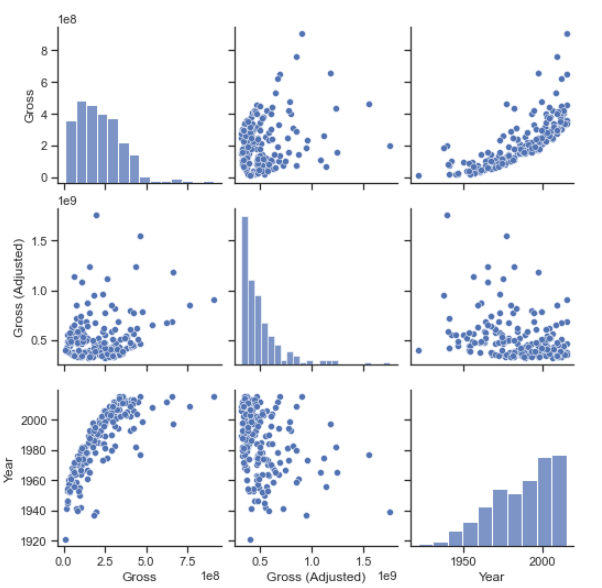




**3.2. For whole dataset, distinguish “Studio” column.**

****

**3.2. For whole dataset, without distinguish “Studio”.**

****