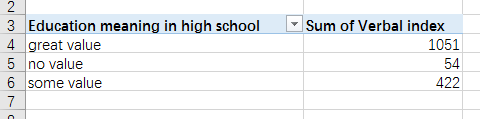
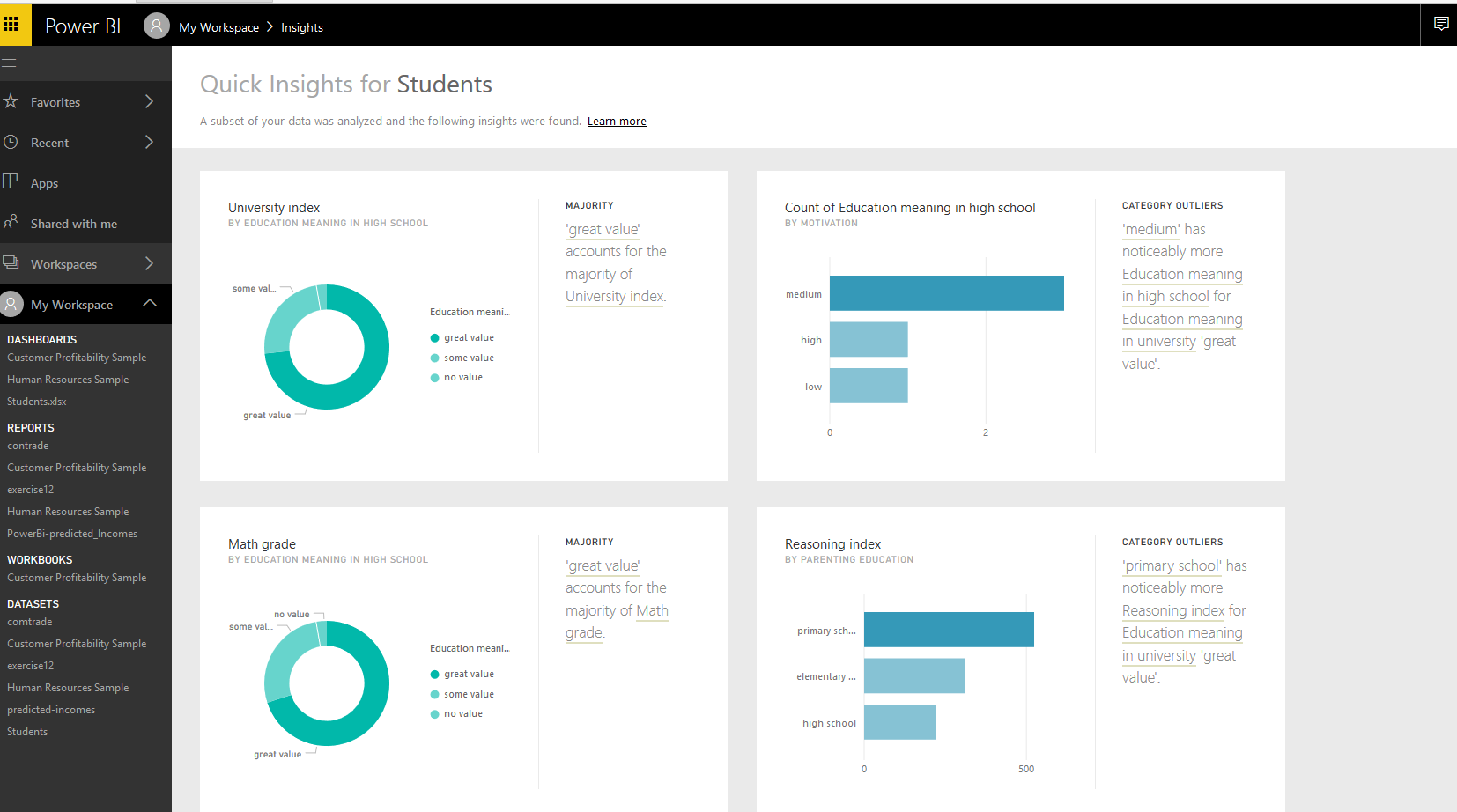
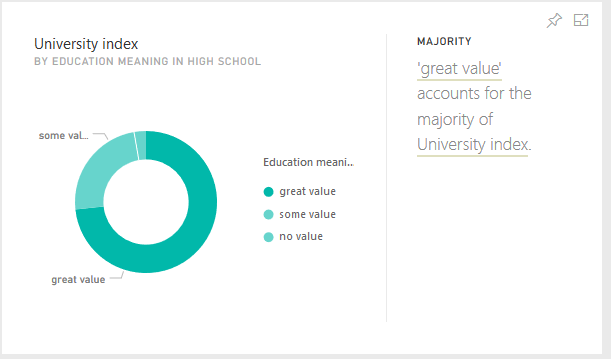
1. Explain charts and tables:

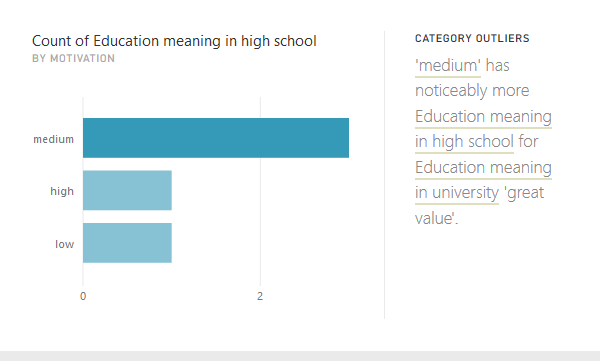


* When I open the file with excel and click right button of the mouse and choose the third one of chart analysis, I got the table and chart, this chart count the sum of verbal index by Education meaning in high school, there are 3 columns in the chart, first column represents students’ sum of verbal index whose education meaning in high school are great value, second column represents students’ sum of verbal index whose education meaning in high school are no value, third column represents students’ sum of verbal index whose education meaning in high school are some value. We can find that students’ sum of verbal index takes the most value when the students’ education meaning in high school is great value. Next is students whose education meaning in high school is some value, the least sum of verbal index happens when students’ education meaning in high school is no value.
* Quick insight: the quick insight gives the analysis of each subset of student.xls, according to this operation.



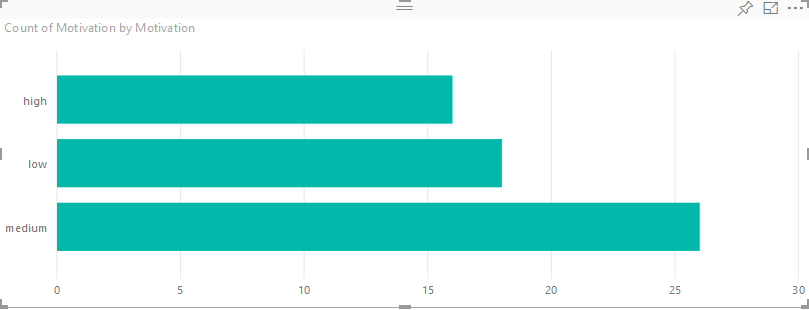


We can see that ‘great value’ accounts for the majority of University index, then is ‘some value’, the lowest is ‘no value’ of education meaning in high school.



We can see that ‘medium’ has noticeably more Education meaning in high school for Education meaning in university ‘great value’, then is the ‘high’ motivation , the lowest is ‘low’ motivation.

1. Make bar charts and illustrate variations within categories and variations within measures

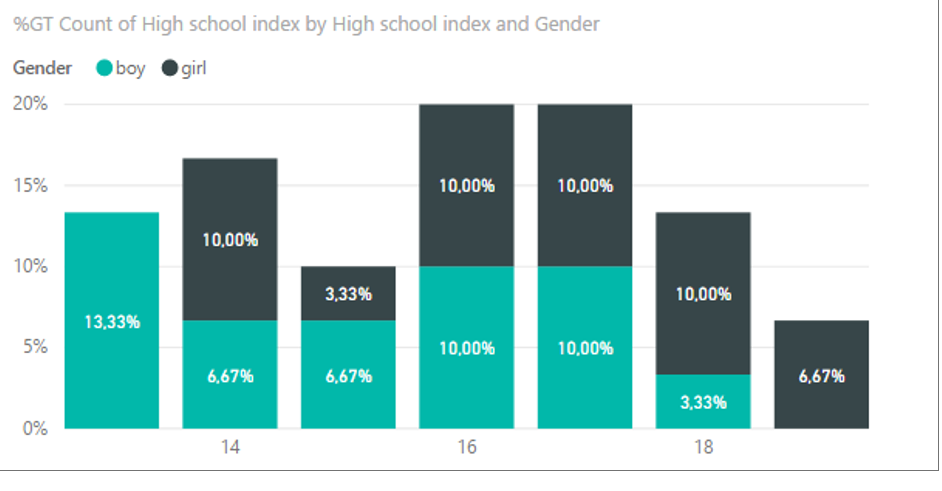


We can see that most students’ motivation is ‘medium’, then is ‘low’, least number of students’ motivation is ‘high’



We can see that when high school indexes are ‘16’ and ‘17’, the count of high school index take the biggest value which is 12, when school index is ‘19’, the count of high school index is the lowest which is ‘4’, which means only 4 students whose school index is ‘19’

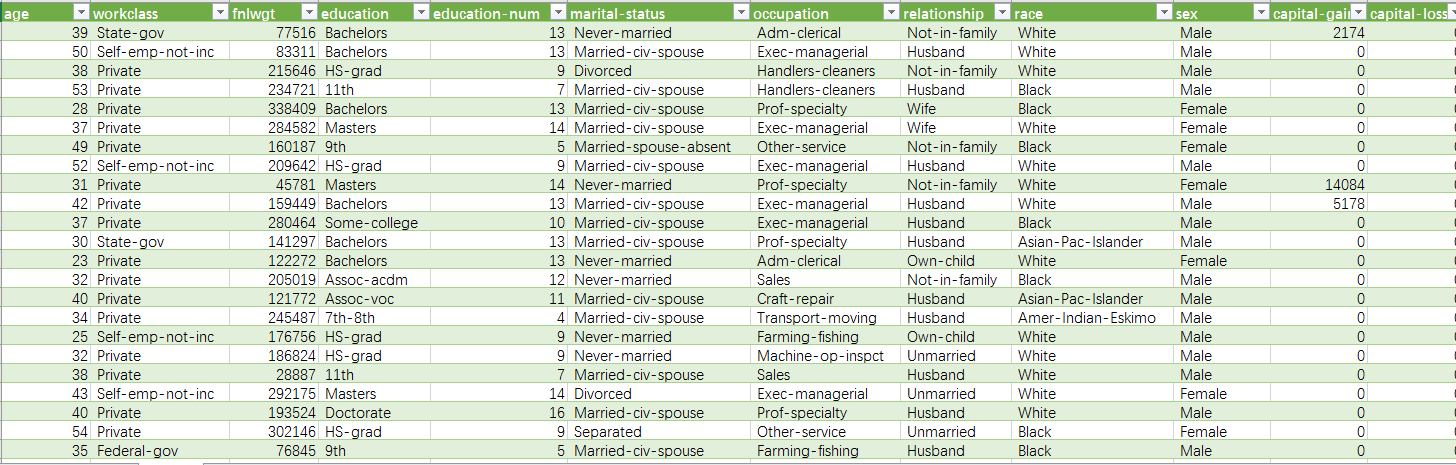
We can see that ‘medium’ motivation takes majority of boys and girls, then is ‘low’ motivation, the least people have ‘high’ motivation; for girls, most of them have ‘medium’ motivation, then is ‘low’ motivation, least of them have ‘high’ and ‘low’ motivation which are both ‘8’; for boys, most of them have ‘medium’ motivation, then are boys with ‘low’ motivation, next is ‘high’ motivation. There are more girls have ‘medium’ motivation than boys, more boys have ‘low’ motivation than girls, same number of boys and girls have ‘high’ motivation.

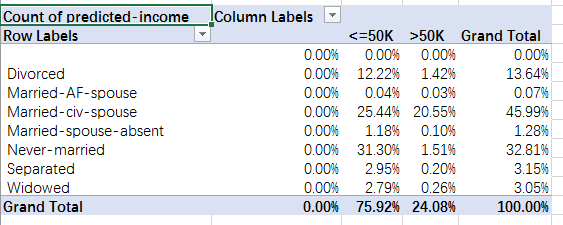


We can see that most of the time, girls’ count of high school indexes are more than boys’, when high school index is ‘13’, there is no girl in this school, when high school index is ‘19’, there is no boy.

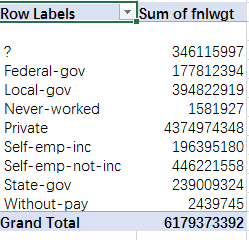
1. Compare the characteristics of MOD and UCI (i.e., similarities and differences)

According to what I have found, MOD is a software platform that provides applicable models that has different target types, number of columns, number of row and so on, users can choose what model they want, and there are docs and links to describe it and give user guidelines of how to use it. Meanwhile, the UCI is a repository that provides users different data sets that differ from default tasks, attribute type, data type, area, instance, format type and so on, users can use those data sets for machine learning and related works. The difference is that MOD provides models and methods, but UCI provides data sets. Their similarity is that they are machine learning oriented, also, users can choose different result by clicking the categories.

1. 5.
2. 
3. (1)



(2)



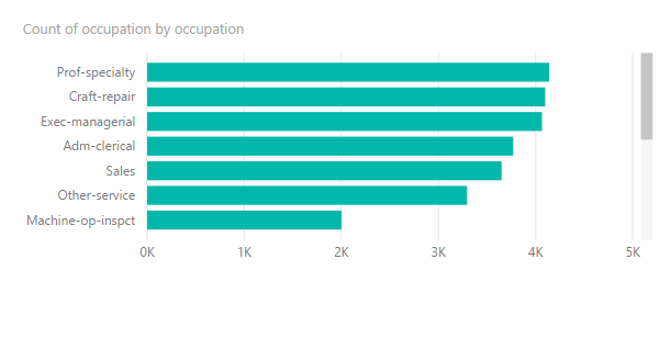
7.8. in the files

9. Explain Median, Variance and Standard deviation:

* Median. Shows the median (middle) value. This is the value that has the same number of items above and below. If there are 2 medians, Power BI averages them.
* Variance: the difference between two values.
* Standard deviation: is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

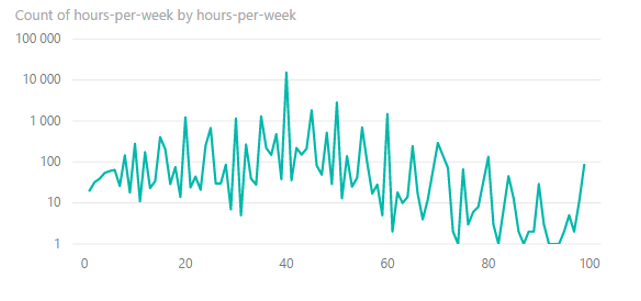
Use the predicted-incomes dataset and illustrate the followings:

* Variation within categories:



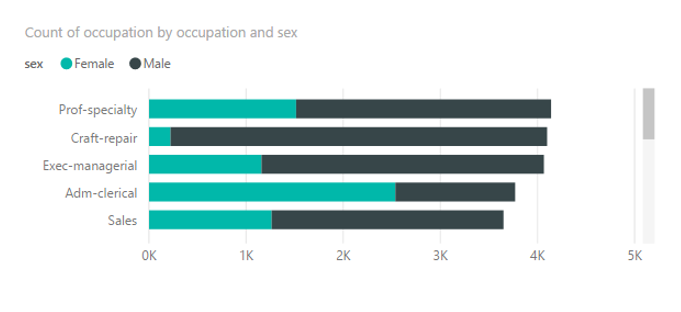
We can see that Prof-specialty takes the majority of occupation, then comes Craft-repair, next is Exec-managerial, Adm-clerical, Sales, other-service, the lowest count is Machine-op-inspct.

* Variation within measures:



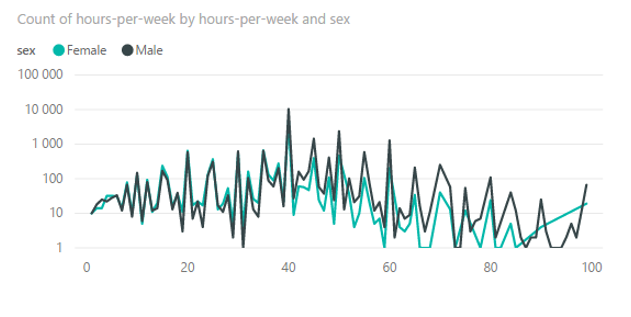
We can see most of the people work 40 hours per week, with the increasing of hours longer than 40, it begins to decrease although there is some heap, when it reaches around 90, there is little people has the value as a working-hour-per-week.

* Relationships among categories:



We can see that Pro-specialty, Craft-repair and Exec-managerial, Sales, for those occupations, there are more males than females, but for Adm-clerical, there are more females than males.

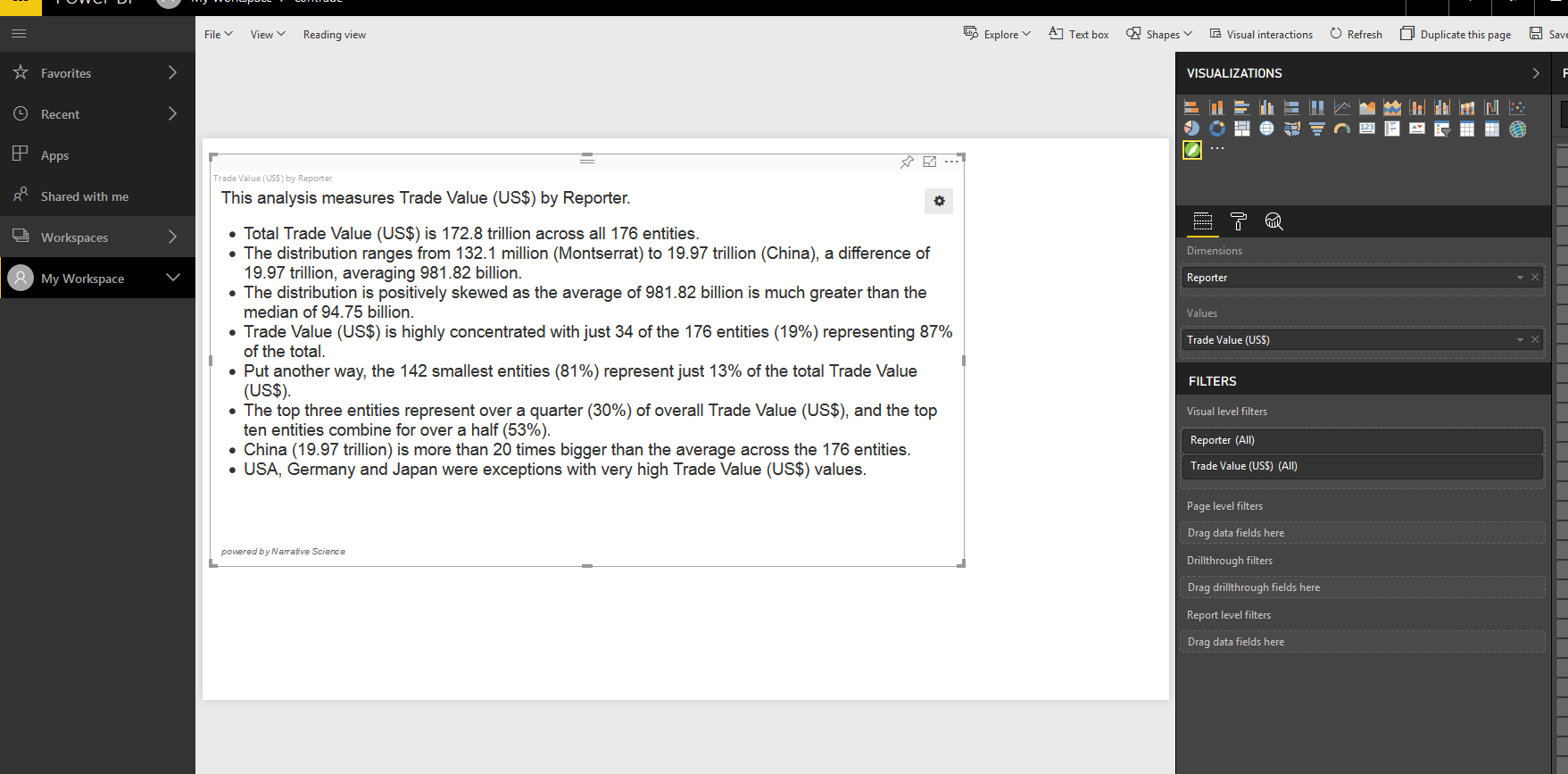
* Relatioships among measures:



We can see that females and males track is almost the same except when the hours-per-week’s value is larger than 60, there are more males works longer than females at that time.

10.

11. 



11,12,13,14,15,16,17,18 in PoweBI

19. top 10: Paper paperboard, Iron and steel, cut, petroleum, petrol product…

Bottom10: Animal, veg, fat, nes, oil, hide, skins, furskins, raw…

20.