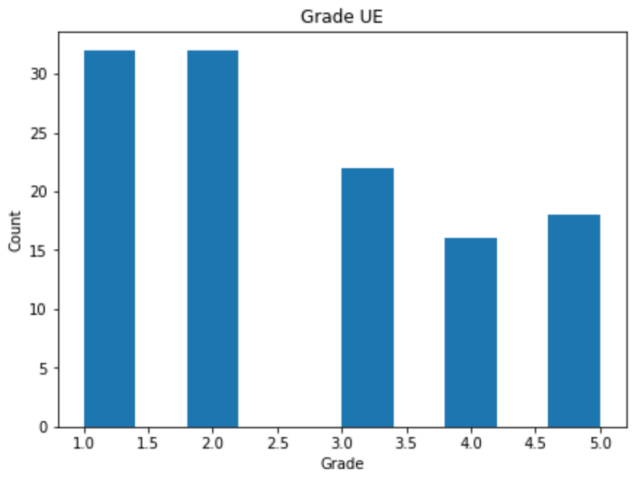
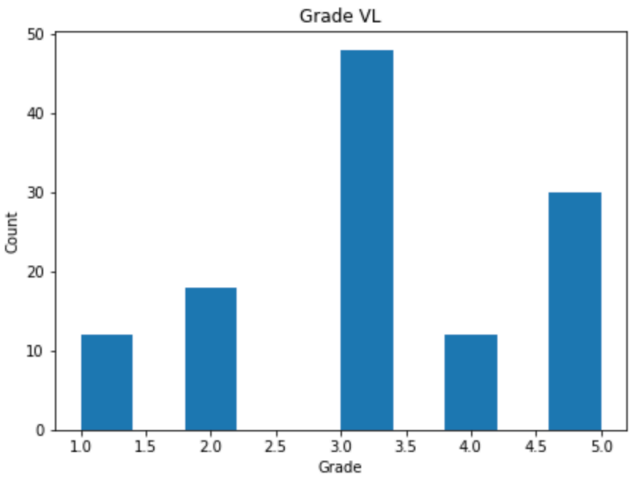
| PROJECT CASE STUDY (258.415/416) | | | SS 2020 |
|---|---|---|---|
| Cur. Id | Student Number | Surname | First name |
| 977 | 11947734 | Luu | Ngoc Tram |

**Task 1: Student Grades**

1. *It's harder to achieve better grades in the UE in general*. Below are frequencies of Grade UE and Grade VL. It's easy to see while the majority of students get grade 1 - 3 in UE class, many students in class can get grade 3 - 5 in VL course. Also, average grade of VL class is also higher than UE.

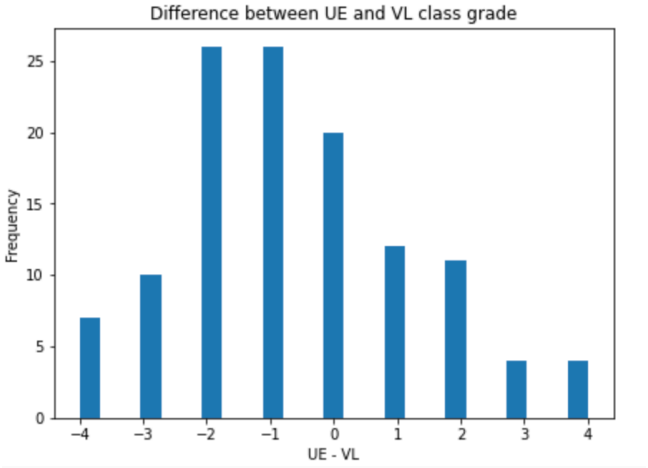



```
from statistics import mean
print('Average grade for UE class is ',mean(ue['Grade UE']))
print('Average grade for VL class is ',mean(vl['Grade VL']))
```

```
Average grade for UE class is  2.6333333333333333
Average grade for VL class is  3.25
```

2. *It's also harder to achieve better grades in the UE for individual students.* To get the result, first is to make a comparison between UE and VL grade of each individual student.

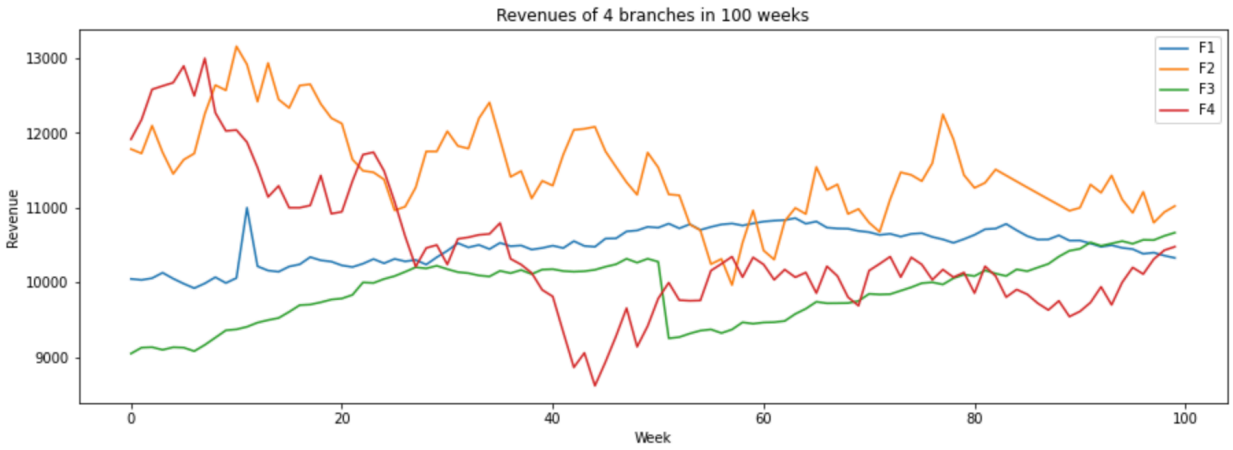



A histogram plot of these comparisons (= UE grade - VL grade) of all students in class is made then. As the histogram is right-skewed, the majority of students got lower UE grade than VL. Most students have their UE grade lower 1 to 2 points compare to VL.

3. To determine whether online attendance has a positive influence on the grades, we look at average grades between students who attend in class and who study online. As average grade of online students is a little bit higher, *it's likely that online study have some positive influence on grades,* even though the difference is not significant.

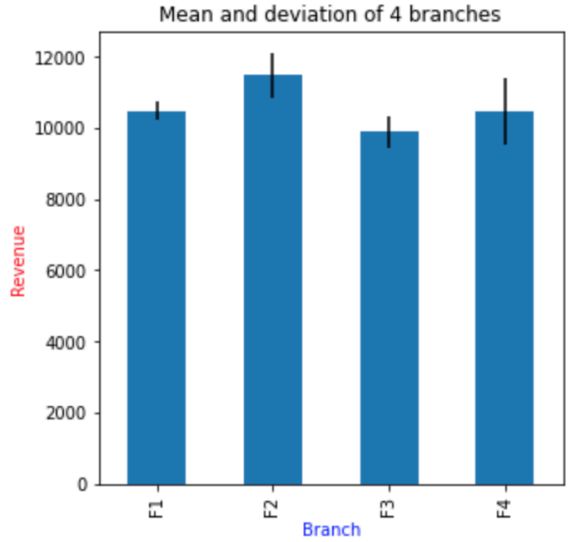
```
AVG.Grade of in class student  2.8676470588235294
AVG.Grade of online student  2.9482758620689653
```

**Task 2: Monthly revenues**




Base on the line graph Revenues of 4 branches in 100 weeks we can observe some trends:

- While revenue of F1 seems to be most stable, revenue of F2 and F4 changed dramatically weeks by weeks. Revenue of F3 grew stable then saw a sharp decline around week 50 and grew up again afterwards.
- In the first 50 weeks the revenues among 4 branches are significantly different, but from week 60 to 100 revenues of all branches tend to be more stable and do not show much difference as previous stage except for branch F2 that have much fluctuated revenue.




Meanwhile, mean and standard deviation in revenues of 4 branches provide some more details:

- While F2 and F4 have the higher average revenue compare with other branches, they also tend to show bigger fluctuation in revenues.
- In particularly F1 is the branch with least fluctuation in revenue, in contrast F4 have the most fluctuated figure.

Base on all the extracted information, we may have some conclusions:

1. F2 and F4 show strong fluctuations in revenues while F1 and F3 are much more stable
2. The differences in revenues between different branches of the company show clearly in the first 50 weeks but from week 60 the differences are not significant. Maybe it's just a recent phenomenon during the first year of operation (~50 weeks) and will end from week 60 when all the activities go into process.

* The appropriate level of measurement for revenues is Cardinal, or more specifically ratio-scaled as it has a meaningful zero-point that be able to multiple to another value and compute the difference between values. For example, revenue = 0 means that in that Quarter that company/ store/… have no revenue and a Finance Analyst can easily sum up revenues of all branches of the company.

**Task 3: Production lines**

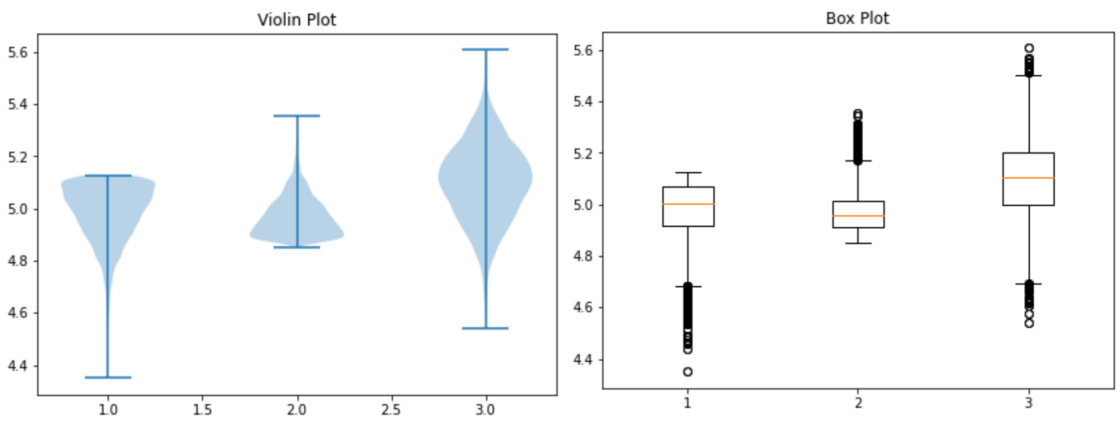
Mean and standard deviation of time making a product of the 3 production lines are as below:

```
Mean produce time of line1:  4.980162690149406
Standard deviation of produce time of line1:  0.11003072553238134

Mean produce time of line2:  4.970096981752859
Standard deviation of produce time of line2:  0.07573801811875343

Mean produce time of line3:  5.0998328652483025
Standard deviation of produce time of line3:  0.15066510195883603
```

1. If we need to choose a production line that can produce as fast as possible, line 2 would be the best choice among all 3 lines as on average it requires the shortest time to produce a product. To produce 1000 products it only needs 4970 seconds, instead of 4980 for line 1 or 5099 for line 3.
2. If we need a production line that can produce just-in-time products, we need to choose a line with the most exact estimation produce time. This means that this line should have smallest differences between time producing each products, also mean that that line will have smallest standard deviation of production time. In this case line 2 would be chosen once again, as standard deviation of production time is just 0.076 instead of 0.11 or 0.15.
3. To have a clearer view, we can also express production time of all 3 lines with box plot or violin plot as below:



| | Violin Plot | Box Plot |
|---|---|---|
| **Usage** | visualize distribution and probability density of the data | only visualize distribution of the data |
| **Information show** | minimum, median, maximum, interquartile range, probability density of the data at different values | minimum, first quartile, median, third quartile, maximum |
| **Advantages** | - more informative as it show the full distribution (distribution and probability) of data<br>- especially useful when data distribution is multimodal<br>- can have multiple layers and format to display variety kinds of information | - more popular with clear figure and easy to understand |
| **Disadvantages** | - unpopular, therefore can be hard for the majority of people to understand | - less informative as it only show distribution of data<br>- can't display clearly characteristics of data when data distribution is multimodal<br>- have fixed format |