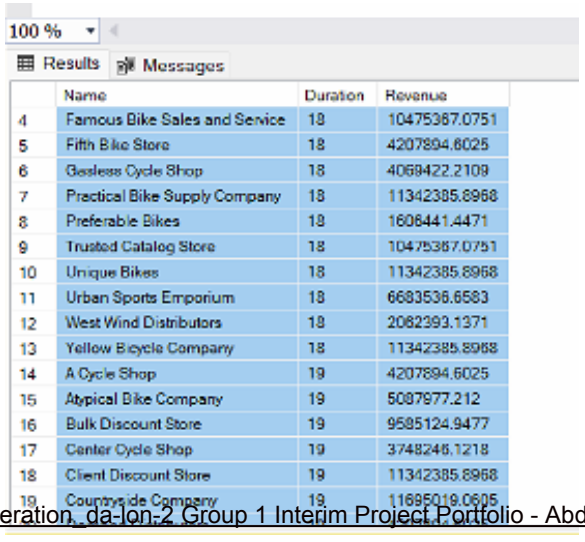
TASK 5 - What is the relationship between store trading duration and revenue?

1. Start using the Microsoft SQL SERVER to explore data. It is useful to check the variables on AdventureWorks_Dictionary.
2. After checking the variabel **Trading Store Duration** and **Revenue**, exploring the AdventureWorks_Dictionary. Was decided to discuss which value(Column) value should We use to align **Trading Store Duration** and **Revenue.**
3. So, We decided to use **TotalDue** placed on **SalesOrderHeader** table as the revenue and **YaerOpened** placed on **View: Sales.vStoreWithDemographics** and subtract the values from last year considering in this database.
4. After It, was created an INNER join from Key_ values placed on Tables SalesOrderHeader (Sale.Person.ID), and from the key values placed on View: Sales.vStoreWithDemographics (BusinessEntity.ID)
5. We also selected the StoreName, Grouped the StoreName and YearOpened and Ordered the result by Duration.
6. So, we've got the query SQL QUERY REVENUE BY TotalDue

```
        SELECT st.Name,
        (2019 - de.YearOpened) AS Duration,
        SUM(soh.TotalDue) AS Revenue
FROM Sales.SalesOrderHeader AS soh
INNER JOIN Sales.Store AS st
        ON soh.SalesPersonID =
st.SalesPersonID
INNER JOIN Sales.vStoreWithDemographics AS de
        ON st.BusinessEntityID =
de.BusinessEntityID
WHERE soh.Status = 5
GROUP BY st.Name, de.YearOpened
ORDER BY Duration;
```

7. After executing,download the Results Question5_table_Revenue_by_totalDue.csv.



	Name	Duration	Revenue
4	Famous Bike Sales and Service	18	10475367.0751
5	Fifth Bike Store	18	4207894.6025
6	Gasless Cycle Shop	18	4069422.2109
7	Practical Bike Supply Company	18	11342385.8968
8	Preferable Bikes	18	1608441.4471
9	Trusted Catalog Store	18	10475367.0751
10	Unique Bikes	18	11342385.8968
11	Urban Sports Emporium	18	6683536.6583
12	West Wind Distributors	18	2082393.1371
13	Yellow Bicycle Company	18	11342385.8968
14	A Cycle Shop	19	4207894.6025
15	Atypical Bike Company	19	5097977.212
16	Bulk Discount Store	19	9585124.9477
17	Center Cycle Shop	19	3748246.1218
18	Client Discount Store	19	11342385.8968
19	Countryside Company	19	11695019.0605

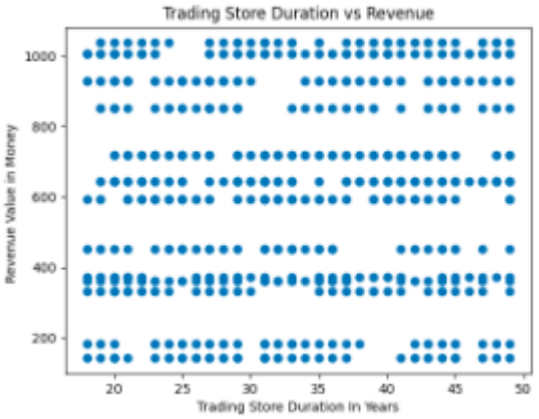
8. Open Python and import the methods pandas, numpy and matplotlib to use functions
   import pandas as pd
   import numpy as np
   from matplotlib import pyplot as plt

9. Use the function pd.read_csv() to run the code.
10. So, to check the relation between Trading Store in duration and Revenue, we used the charts : **Scatterplot**, **Line plot** because they are all used to compare continuous variables and check the correlation between theirs. We also did a bar chart and histogram for Revenue values (to check how the distribution was). We used some function to name the axis, add title, and divide the values.
We saved the code on Question5_Revenue_Total_Due.py

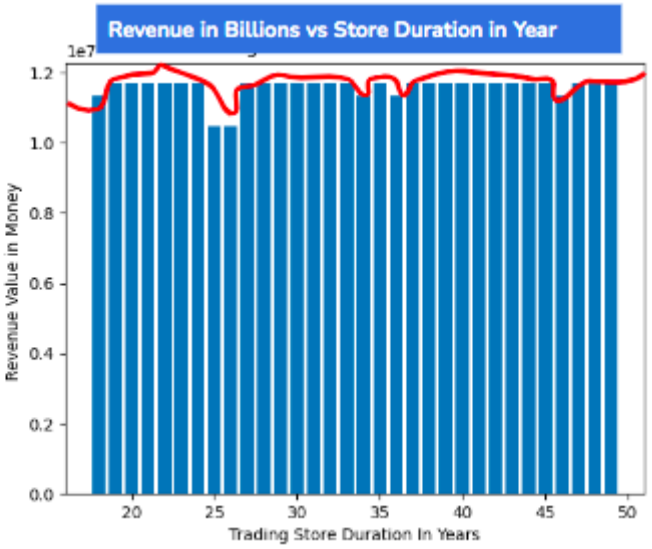
```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt

#td means trading duration vs revenue
td = pd.read_csv("datasets/Question5_table_Revenue_by_totalDue.csv")
print(td.head())

plt.scatter(td.Duration, td.Revenue)
plt.title("Trading Store Duration vs Revenue in Bilions$")
plt.xlabel("Trading Store Duration In Years")
plt.ylabel("Revenue Value in Money ")
plt.show()

#What is the relationship between Trading Store Duration and Revenue ?
#plt.bar(td.Duration, td.Revenue)
#plt.barh (td.Duration, td.Revenue)
#plt.hist (td.Duration)
#plt.hist (td.Duration)
#plt.title("Trading Store Duration vs Revenue")
#plt.xlabel("Trading Store Duration In Years")
#plt.ylabel("Revenue Value in Money ")
#plt.show()
```

11.After executing the charts below, we have got the conclusion that there's no relation between Trading Store Duration and Revenue. As we can see on the charts below. Because the scatter chart comes with points not relationed, and also, checking the bar chart, we could see a small variation which is considered not relevant.

There's none Relation between Store Trading Duration (in Years) and Revenue. The Revenue from each store name are not relation with your duration time.



We can see the revenue decreased in specific duration stores, but we can not consider it as relevant). There's none variation.



Also, the Revenue value there's none normalized distribution. Which doesn't mean much, but assumes he doesn't follow a standard behavior