**Cleaning data of supply chain**

**First: using Python**

* **A screenshot of a computer

  AI-generated content may be incorrect.Read file in python**
* **A screenshot of a computer

  AI-generated content may be incorrect.Checking missing values:**

**There are no missing values in all variables in dataset**

**Also, the data types of all variables are suitable for the data inserted into each one of them**

* **A screenshot of a computer

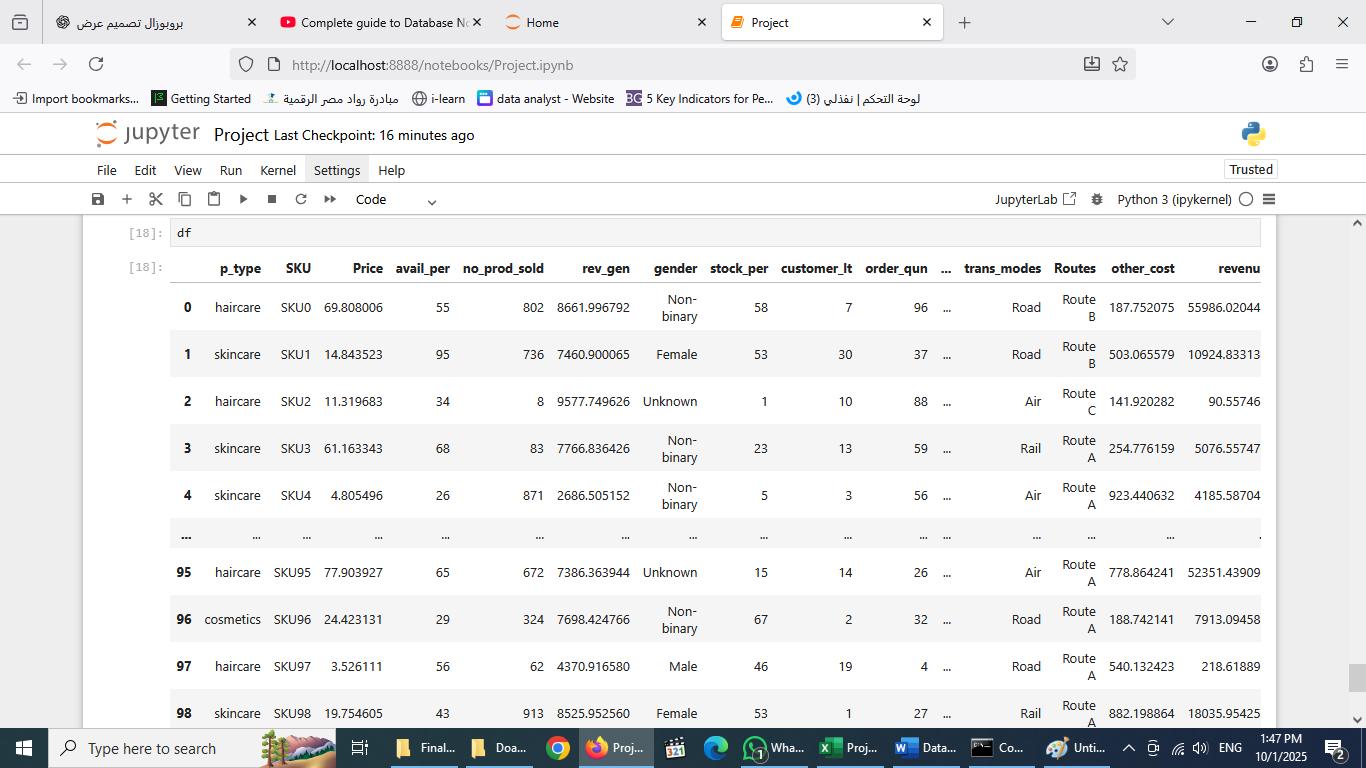
  AI-generated content may be incorrect.Checking duplicated values:**

**There are no duplicated values in dataset**

* **Renaming variables:**

A screenshot of a computer

AI-generated content may be incorrect.To avoid ambiguous columns names, I renamed many columns



**Second: using SQL**

* **Checking missing values:**

There are no missing values, so, the result of running this query for all columns (25) is no record

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

* **Checking duplicated values**

There are no duplicated values, the result of running this query is no records

A computer screen with text boxes

AI-generated content may be incorrect.

* **Renaming column:**

We use this query to rename all columns with ambiguous names. There are 3 methods we can use to rename column in SQL:

1. When importing file to SQL, we can rename columns in importing steps, or
2. We can right click on column name and choose modify then rename it, or
3. ask ChatGPT and we get this command (renaming) in SQL\_Server.

exec sp\_rename 'project.availablity', 'avail\_per'

**Third: using Excel**

* **checking missing values:**

we can do this by two ways:

1. A screenshot of a computer

   AI-generated content may be incorrect.using power query
2. A screenshot of a computer

   AI-generated content may be incorrect.or by using conditional formatting to highlight blanks and then remove it:

* **Checking duplicated rows:**

Also, can be done using three ways:

1. A screenshot of a computer

   AI-generated content may be incorrect.Using power query
2. A screenshot of a computer

   AI-generated content may be incorrect.Or using conditional formatting to highlight duplicated rows and then remove it.
3. A screenshot of a computer

   AI-generated content may be incorrect.Or choose (remove duplicates) from menu after selecting the whole table

* **Rename columns:**

Also, there are two ways:

1. A screenshot of a computer

   AI-generated content may be incorrect.Using power query:
2. A screenshot of a computer

   AI-generated content may be incorrect.In sheet of data, I can just select the column name and change it:

**The variables names after cleaning**

|  |  |  |  |
| --- | --- | --- | --- |
| **Index** | **Original Names** | **Understanding Names**  **(for explanation)** | **Names for analysis (shortcuts with no spaces – with no special characters)** |
| **1** | **Product type** | **Category** | **p\_type** |
| **2** | **SKU** | **product ID** | **SKU** |
| **3** | **Price** | **Price** | **Price** |
| **4** | **Number of products sold** | **No of products sold** | **no\_prod\_sold** |
| **5** | **Revenue generated** | **Revenue generated** | **rev\_gen** |
| **6** | **Availability** | **Availability %** | **avail\_per** |
| **7** | **Customer demographics** | **Gender** | **gender** |
| **8** | **Stock Level** | **Stock Level %** | **stock\_per** |
| **9** | **Lead times** | **Customer Lead time** | **customer\_lt** |
| **10** | **Order quantities** | **No of Orders** | **order\_qun** |
| **11** | **Shipping times** | **Shipping times** | **shipping\_lt** |
| **12** | **Shipping carriers** | **Shipping carriers** | **shipping\_carr** |
| **13** | **Shipping costs** | **Shipping costs** | **shipping\_cost** |
| **14** | **Supplier name** | **Supplier name** | **sup\_name** |
| **15** | **Location** | **Location** | **delv\_location** |
| **16** | **Lead time** | **Supplier Lead time** | **supplier\_lt** |
| **17** | **Production volumes** | **Production volumes** | **prod\_volum** |
| **18** | **Manufacturing lead time** | **Manufacturing lead time** | **manfuct\_lt** |
| **19** | **Manufacturing costs** | **Manufacturing costs** | **manfuct\_cost** |
| **20** | **Inspection results** | **Inspection results** | **inspect\_resl** |
| **21** | **Defect rates** | **Defect rates** | **def\_rate** |
| **22** | **Transportation modes** | **Transportation modes** | **trans\_modes** |
| **23** | **Routes** | **Routes** | **Routes** |
| **24** | **Costs** | **Other Costs** | **other\_cost** |