CASE STUDY

You are a Data Analyst in a manufacturing company, All-rounder. All-rounder manufactures a variety of products from Furniture to Office supplies to Technology.

Mr. Bid, the Head of Sales department, has given you a little task. He has asked you to mine the company's sales data containing sales information from year 2014 – 2018 and gather some insights that he needs in the next 5 hours.

Things to note about this project;

- The formula for calculating Total Average Revenue per Customer (product) is dividing the Total Revenue by the Total number of customers (products).
- The Year-Over-Year growth rate can be computed using either of these formulas: (Total Revenue from the current year / Total Revenue from the previous year) - 1, or (Total Revenue from the current year - Total Revenue from the previous year) divided by Total Revenue from the previous year.
- To calculate the Cost of Goods Sold (COGS), subtract the Revenue from the Profit. In this context, Revenue is referred to as Sales Revenue.
- The "Ship_date" column can be used to extract the year values.
- When working with the data, avoid making any changes to the original table, except for data cleaning.
- Some of the tables provided might be incomplete, but they are included to give an idea of what the expected result should look like.
- The dataset used for this task is a superstore data retrieved from kaggle.com.

That Being Said, let's begin...

- After importing the data, write your select all statement to know what your data looks like. Then clean the data.

1. Find out the Total Average Revenue per Customer and the Total Average Revenue per Product for the year 2017. (Return values in 2 decimal places).

Avg_rev_per_customer	Avg_rev_per_product
1057.73	479.27

2. Give a list of all products that generated revenue below the TAR_Product (Showing the le ast revenues first).

product_id	product_name	Revenue
OFF-AP-10001634	Hoover Commercial Lightweight Upright Vacuum	1.39
OFF-BI-10003094	Self-Adhesive Ring Binder Labels	1.41
OFF-AP-10002203	Eureka Disposable Bags for Sanitaire Vibra Groomer I Upright Vac	1.62
OFF-BI-10002609	Avery Hidden Tab Dividers for Binding Systems	1.79
OFF-FA-10002676	Colored Push Pins	1.81
OFF-FA-10000840	OIC Thumb-Tacks	1.82

2b. Find out the number of products that fall below the TAR/P under each category per region.

Region	Office_Supplies	Furniture	Technology
South	116	46	38
East	274	94	92
Central	258	72	46
West	330	126	125

3. Find out the Y-O-Y growth rate from 2014 – 2017. (Your answer should be rounded to 3 decimal places, and show the "%" sign).

YOY_2015	YOY_2016	YOY_2017
1.926%	27.508%	19.558%

4. Return a table that shows the Total Revenue against the Total Cost of goods sold (COGS) for each region and in each year (2014 – 2017).

Year	Rev_East	COGS_East	Rev_West	COGS_West	Rev_South	COGS_South
2016	180090.4	159947.3	190144.1	165166.5	94019.88	76047.06
2015	164991.7	142133	139903.7	119879.8	71324.55	63097.16
2014	119992.2	104710.9	144206.3	124315.3	103284.8	91592.85
2017	213105.5	179957.1	249425.2	206205.3	121213.2	112837.8

5. Return a table that shows the Profit generated per state, then per region in 2018 alone.

State	South	East	Central	West
California	0.0	0.0	0.0	237.3036
New York	0.0	91.0702	0.0	0.0
Indiana	0.0	0.0	56.511	0.0
Texas	0.0	0.0	-125.869	0.0
Iowa	0.0	0.0	20.585	0.0
Washington	0.0	0.0	0.0	33.176
Michigan	0.0	0.0	100.656	0.0

6. Return a table that shows the total number of orders received on each day of the week for year 2014, 2015, 2016 and 2017

Day_Of_Week	Order_2014	Order_2015	Order_2016	Order_2017
Sunday	291	394	461	564
Monday	404	359	461	647
Tuesday	352	199	285	270
Wednesday	210	34	46	81
Thursday	50	388	479	546
Friday	359	393	461	605
Saturday	327	335	394	599

Note that this challenge was gotten from https://github.com/AkDeAnalyst