iNeuron

Low Level Design

Analyzing Amazon Sales Data

Written by	Szimonetta Farkas
Document Version	0.1
Last Revised Date	04/08/2023

DOCUMENT CONTROL

VERSION	DATE	AUTHOR	COMMENTS
0.1	04/08/2023	Szimonetta Farkas	First version of complete LLD

iNeuron

Contents

1. Introduction

1.1 What is Low-Level design document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Analyzing Amazon Sales Data Project Dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2. Architecture

2.1. POWER BI

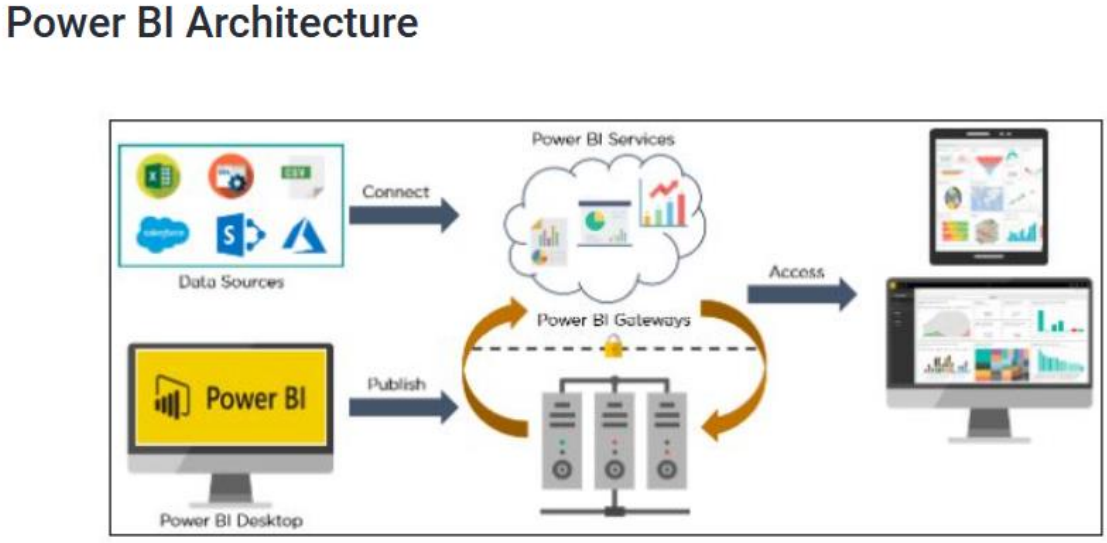


Photo: (1)

Power BI is a business intelligence platform created by Microsoft. It is used for business analytics, data visualization and it presents actionable information from raw data.

Power BI has a user-friendly surface, and it is very helpful with data integration. Data can be integrated from various sources: excel, text/csv, xml, json, web files, SQL Server, cloud-based sources (Azure, Salesforce).

With the help of Power BI we can create customizable dashboards and we can view up-to-date data because it supports real-time data processing. Power BI also gives the opportunity to collaborate and work on data analysis models project together.

Power BI has some disadvantages as well: it has limited processing capabilities, limited customization options and it is not a free tool, users have to pay for additional features and for storage space.

Power BI technology has more components such as: Power Query, Power Pivot, Power View, Power Map, Power BI Desktop, Power BI Mobile, Power Q&A (1).

2.2. Python

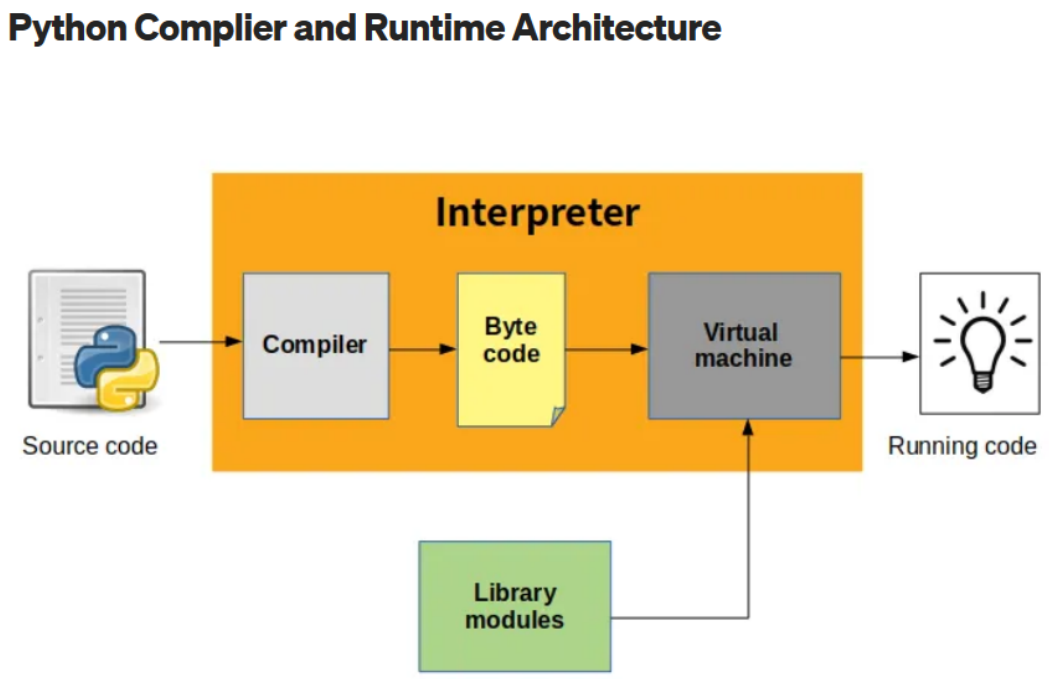


Photo: (2)

Python is an open-source, object-oriented interpreted programming language (2). It works on different platforms such as Windows, Mac, Raspberry, Linux, etc. Python programming language is easily readable and it has a simple syntax. It allows programmers to write programs with fewer lines than other programming languages (3).

Python is used for analyzing data, creating web applications, creating workflows, connecting to databases, handling big data, performing complex mathematical tasks, rapid prototyping and software development (3).

Most important Python libraries: Pandas, NumPy, Keras, TensorFlow, Scikit Learn, Eli5, SciPy, PyTorch, LightGBM, Theano (4).

IDE for Python: Pycharm by Jetbrains, Visual Studio Code by Microsoft.

Notebooks for Python: Jupyter Notebook, Google Colab, Databricks (2).

3. Architecture Description

3.1. Data Description

The dataset is in excel format. It contains sales information about customer orders from the year 2017, 2018 and 2019.

Amazon Sales Data features:

CustKey: unique ID to identify customers

DateKey: transaction date

Discount Amount: difference between sales amount based on list price and sales amount

Invoice Date: date on which the invoice created

Invoice Number: unique number generated after the invoice

Item Class: class of the item

Item Number: unique number to identify the item

Item: name of item

Line Number: number of line from which it is ordered

List Price: price quoted by by the manufacturer

Order Number: unique ID to identify the order

Promised Delivery Date: date when the delivery is expected

Sales Amount: sales prices * quantity

Sales Amount based on List Price: list price * quantity

Sales Cost Amount: amount caused for making sales of the item

Sales Margin Amount: sales amount – sales cost amount

Sales Price: price at which the item is sold

Sales Quantity: quantity of the item ordered

Sales Rep: unique ID for the sales representative

U/M: unit of measurement

3.2. Loading the Dataset

With the help of the Pandas library, I loaded the raw dataset to Python and I discovered it.

Loading The Dataset

```
data = pd.read_excel(r'E:\Data_Science\INTERNSHIP\iNeuron_Amazon_Sales\SALESDATA.xlsx',
                     parse_dates=['DateKey', 'Invoice Date', 'Promised Delivery Date']) # creating a dataframe
```

Discovering The Dataset

```
data.head(10) # first 10 rows
```

	CustKey	DateKey	Discount Amount	Invoice Date	Invoice Number	Item Class	Item Number	Item	Line Number	List Price	...	Sales Amount	Sales Amount Based on List Price	Sales Cost Amount	Sales Margin Amount	Sales Price	S Qua
0	10000481	2017-04-30	-237.910	2017-04-30	100012	NaN	NaN	Urban Large Eggs	2000	0.000	...	237.91	0.000	0.0	237.91	237.910000	
1	10002220	2017-07-14	368.790	2017-07-14	100233	P01	20910	Moms Sliced Turkey	1000	824.960	...	456.17	824.960	0.0	456.17	456.170000	
2	10002220	2017-10-17	109.730	2017-10-17	116165	P01	38076	Cutting Edge Foot-Long Hot Dogs	1000	548.660	...	438.93	548.660	0.0	438.93	438.930000	
3	10002489	2017-06-03	-211.750	2017-06-03	100096	NaN	NaN	Kiwi Lox	1000	0.000	...	211.75	0.000	0.0	211.75	211.750000	
4	10004516	2017-05-27	96627.940	2017-05-27	103341	P01	60776	High Top Sweet Onion	1000	408.520	...	89248.66	185876.600	0.0	89248.66	196.150901	

3.3. Preparing the Dataset for Analysis

I removed the null values from the dataset, made it ready for analysis in Power BI.

```
# removing null values from the dataset
data1.dropna(subset=['Discount Amount', 'Item Number', 'Sales Price'], inplace = True)
```

```
data1.isna().sum()
```

```
CustKey                          0
DateKey                          0
Discount Amount                  0
Invoice Date                     0
Invoice Number                   0
Item Class                       0
Item Number                      0
Item                             0
Line Number                      0
List Price                       0
Order Number                     0
Promised Delivery Date           0
Sales Amount                     0
Sales Amount Based on List Price 0
Sales Cost Amount                0
Sales Margin Amount              0
Sales Price                      0
Sales Quantity                   0
Sales Rep                        0
U/M                              0
Year                             0
Month                            0
Quarter                          0
Day                              0
dtype: int64
```

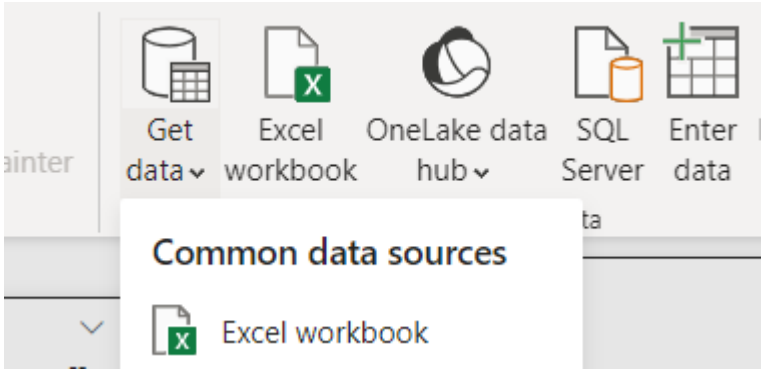
3.4. Creating new columns.

I created new columns which were required for the further analysis. I saved the cleaned dataset as an excel file.

```
# creating new columns: year, month, quarter, day
data1['Year'] = data1['Invoice Date'].dt.year
data1['Month'] = data1['Invoice Date'].dt.month
data1['Quarter'] = data1['Invoice Date'].dt.quarter
data1['Day'] = data1['Invoice Date'].dt.day
```

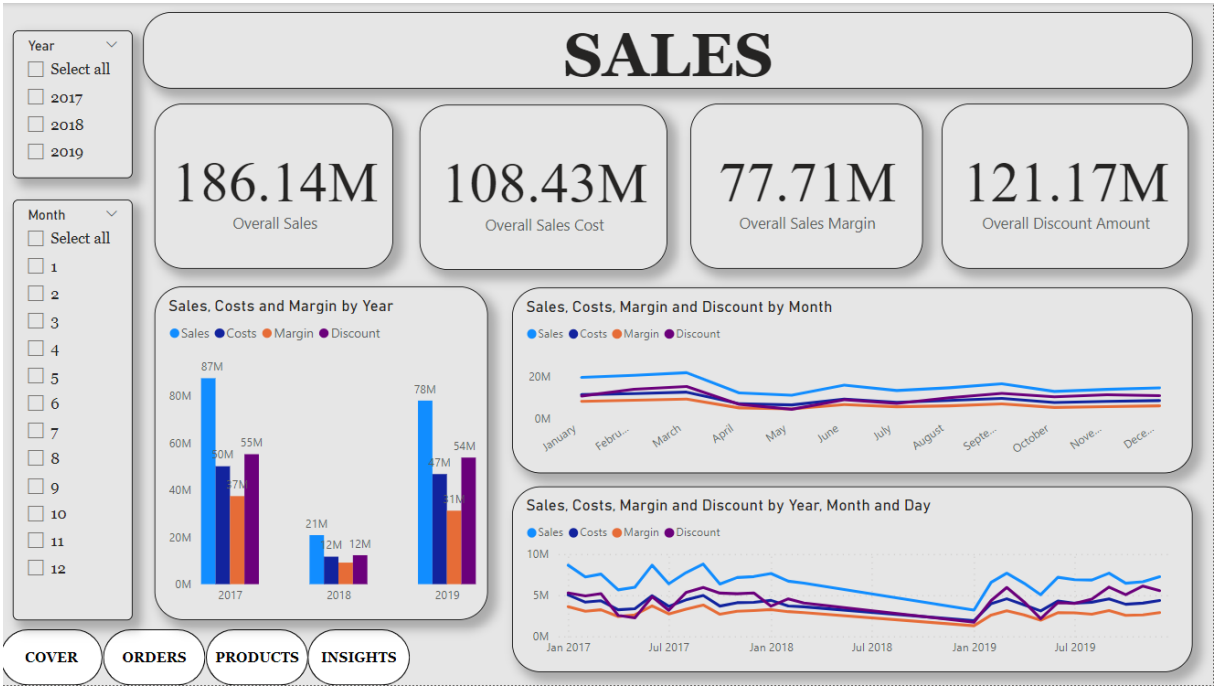
3.5. Loading the Cleaned Dataset into POWER BI

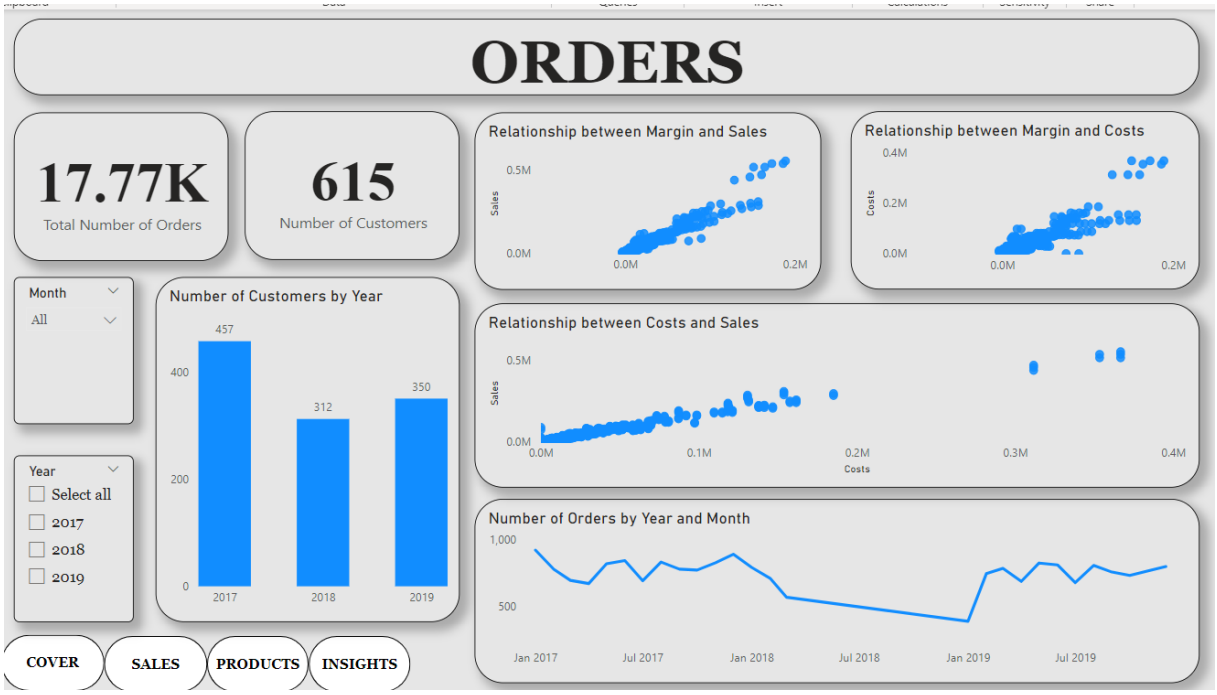
I opened the Power BI Desktop, clicked on: Get Data → Excel Workbook. I browsed the file and I clicked on Transform Data. As my dataset was cleaned, I just took a look at it in the Power Query Editor, then I started the visualization.

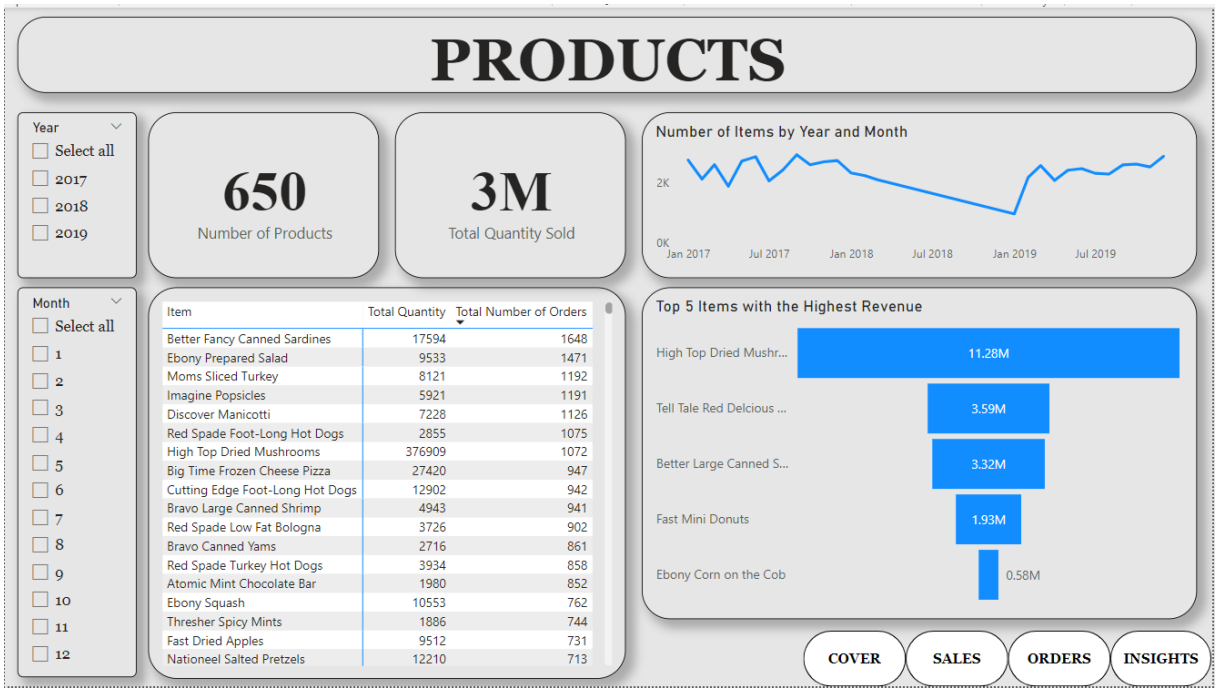


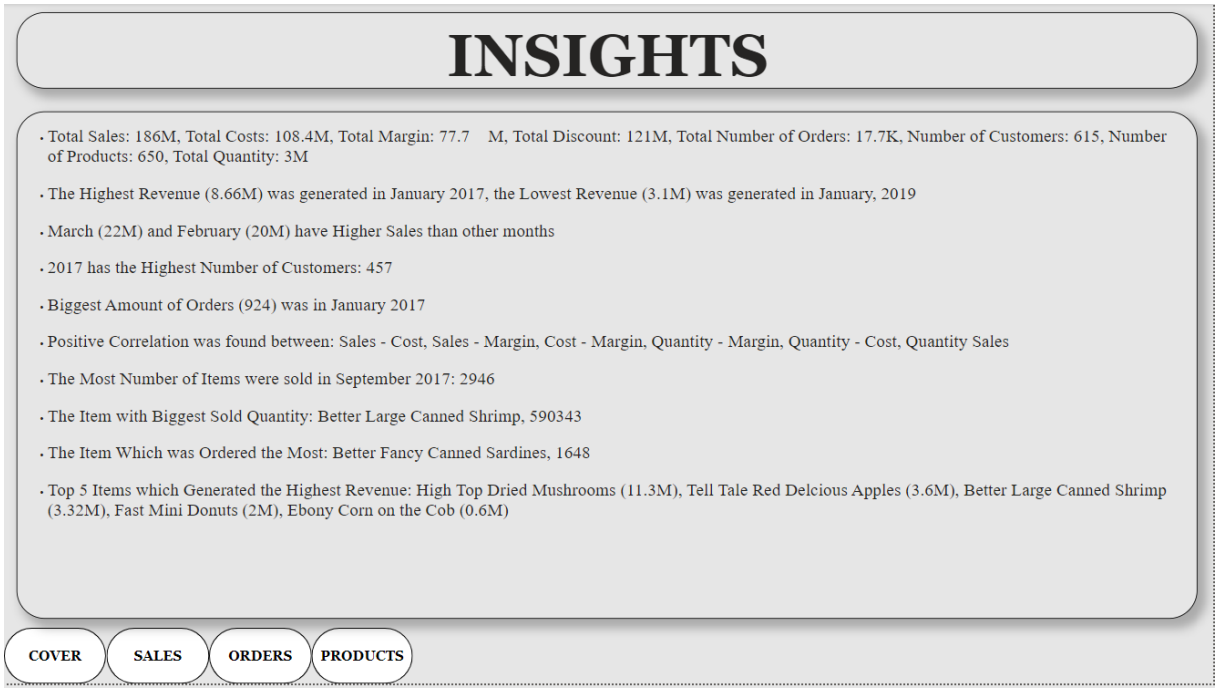
3.6. Creating Dashboard

In the Report View I created the visualization and found the answer for the business questions. I created 5 pages: Cover, Sales, Orders, Products, Insights. I do not have access to Power BI Services, so I did not publish my dashboard.




ORDERS
17.77K
Total Number of Orders
615
Number of Customers
Relationship between Margin and Sales
Relationship between Margin and Costs
Month
All
Year
Select all
2017
2018
2019
Number of Customers by Year
457
312
350
2017
2018
2019
Relationship between Costs and Sales
Number of Orders by Year and Month
COVER
SALES
PRODUCTS
INSIGHTS


PRODUCTS
Year
Select all
2017
2018
2019
650
Number of Products
3M
Total Quantity Sold
Number of Items by Year and Month
Month
Select all
1
2
3
4
5
6
7
8
9
10
11
12
Item
Total Quantity
Total Number of Orders
Better Fancy Canned Sardines 17594 1648
Ebony Prepared Salad 9533 1471
Moms Sliced Turkey 8121 1192
Imagine Popsicles 5921 1191
Discover Manicotti 7228 1126
Red Spade Foot-Long Hot Dogs 2855 1075
High Top Dried Mushrooms 376909 1072
Big Time Frozen Cheese Pizza 27420 947
Cutting Edge Foot-Long Hot Dogs 12902 942
Bravo Large Canned Shrimp 4943 941
Red Spade Low Fat Bologna 3726 902
Bravo Canned Yams 2716 861
Red Spade Turkey Hot Dogs 3934 858
Atomic Mint Chocolate Bar 1980 852
Ebony Squash 10553 762
Thresher Spicy Mints 1886 744
Fast Dried Apples 9512 731
Nationeel Salted Pretzels 12210 713
Top 5 Items with the Highest Revenue
High Top Dried Mushr...
11.28M
Tell Tale Red Delcious ...
3.59M
Better Large Canned S...
3.32M
Fast Mini Donuts
1.93M
Ebony Corn on the Cob
0.58M
COVER
SALES
ORDERS
INSIGHTS



4. Unit Test Cases

Test Case Description	Expected Results
Slicer of Year, Month	Work Properly
Charts	Work Properly
Page Buttons	Work Properly

5. References

1. What is Power BI?: Services, Architecture, and Dashboard [Updated] | Simplilearn
2. #Series 1 Python Basic Architecture, Installation, IDE's, Indentation & Comments | by Keren Melinda | Medium
3. Introduction to Python (w3schools.com)
4. List of Top 10 Libraries in Python (2023) - InterviewBit