Low Level Design (LLD)

Amazon Sales Analysis

Revision number 1.2
Last Date of Revision–21/09/2022

Aishwarya Poman

Document Control

Date	Version	Description	Author
4/06/2022	1.0	Introduction, Problem Statement	Aishwarya Poman
24/08/2022	1.1	Dataset Information, Architecture Description	Aishwarya Poman
21/09/2022	1.2	Final Revision	Aishwarya Poman

Why this Low-Level Design Document?

The purpose of this document is to present a detailed description of the Sales analysis technique. It will explain the necessary steps which have to be followed before any analysis can begin. LLD 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. This document is intended for both the stakeholders and the developers of the system and will be proposed to the higher management for its approval.

The LLD will be focusing on the below objectives:

- Problem Understanding.
- Data Acquisition.
- Data Pre-Processing and Exploratory Analysis
- Auditing accuracy
- Dashboard report for important activities

Scope

The LLD documentation presents the detailed structure of the Investment analytics for each of its individual components. The goal of LLD is to give the internal logical design of the actual program code. Low-level design is created based on the high-level design.

Project Introduction

Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise. So we are performing the sales analysis for amazon food sales.

Constraints

Our analysis is done based on a limited dataset provided for 3 years and different food items. The analysis is done year wise.

Risks

Document specific risks that have been identified or that should be considered.

Out of Scope

Delineate specific activities, capabilities, and items that are out of scope for the project.

1. Technical specifications:

Dataset

The Dataset is taken from iNeuron's provided dataset-

```
In [4]: pd.read_excel('SALESDATA.xls')
```

Out[4]:

	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
0	10010923	2018-03-21	1605.9000	2018-03-21	226563	P01	63560	Gorilla Low Fat String Cheese	8000	298.0000	...	1374.10	2980.0000	954.13	419.97	137.410000
1	10010923	2018-03-21	2408.8500	2018-03-21	226563	P01	62616	Tell Tale Limes	10000	298.0000	...	2061.15	4470.0000	1430.19	630.96	137.410000
2	10010923	2018-03-21	603.9900	2018-03-21	226563	P01	28929	Nationeel Potato Chips	1000	37.3600	...	516.81	1120.8000	222.82	293.99	17.227000
3	10010923	2018-03-21	272.1710	2018-03-21	226563	P01	61484	Super Creamy Peanut Butter	4000	50.5051	...	232.88	505.0510	152.43	80.45	23.288000
4	10010923	2018-03-21	481.7700	2018-03-21	226563	P01	63559	Gorilla String Cheese	9000	298.0000	...	412.23	894.0000	285.44	126.79	137.410000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
65277	10002969	2019-06-08	173.3500	2019-06-08	303217	P01	45880	Red Spade Low Fat Bologna	3000	433.3800	...	4160.45	4333.8000	2807.49	1352.96	416.045000
65278	10015495	2019-06-28	-333.7000	2019-06-28	305106	NaN	36013	Big Time Frozen Mushroom Pizza	1000	0.0000	...	333.70	0.0000	195.05	138.65	8.342500
[illegible]	[illegible]	2019-05-	[illegible]	2019-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	...	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

```
In [5]: pd.read_excel('CUSTOMERADDRESS.xls')
```

Out[5]:

	Address Number	City	Country	Customer Address 1	Customer Address 2	Customer Address 3	Customer Address 4	State	Zip Code
0	10000471	Jeddah	AU	Al Thalia Street	PO Box 11605			NaN	11443
1	10004255	Seoul	AU	3F Shinwoo Building	195-12 Poi-Dong Kangnam-KU			NaN	135-260
2	10007117	Taipei	AU	8F No.19 Chang Chun Road				NaN	104
3	10010923	Seocho-GU	AU	231 Yangjae Dong				NaN	137-938
4	10013011	112 Reyjavik	AU	Vagnhofda 23				NaN	
...	...	...	...	...	...	...	...	...	...
450	10027560	Odessa	US	3356 Kermit Highway				TX	79764
451	10027572	Elma	US	2210 Bowen Road				NY	14059
452	10027575	Dallas	US	10400 Plano Road				TX	75238
453	10027583	Morton	US	Attention: Charlene Hoyer	500 North Morton Avenue	PO Box 474		IL	61550-0474
454	10027629	Dearborn	US	Central Accounting Service	PO Box 6005			MI	48121

455 rows × 9 columns

```
In [6]: pd.read_excel('CUSTOMERS.xls')
```

Out[6]:

	Address Number	Business Family	Business Unit	Customer	Customer Number	Customer Type	Division	Line of Business	Phone	Region Code	Regional Sales Mgr	Search Type	Unnamed: 12
0	10000000	R3	1	City Supermarket	10000000	G2	2		816-455-8733	4	S16	C	NaN
1	10000453	R3	1	A Supermarket	10000453	G1	1		816-455-8733	5	S19	C	NaN
2	10000455	R3	1	Caribian Supermarket	10000455	G2	2		816-455-8733	1	S16	C	NaN
3	10000456	R1	1	A&B Shop	10000456	G3	1		816-455-8733	0	S2	C	NaN
4	10000457	O2	1	A&G Shop	10000457	G1	1		816-455-8733	5	S1	C	division
...	...	...	...	...	...	...	...	...	...	...	...	...	...
679	10027560	R2	1	Zilog Shop	10027560	G2	2		816-455-8733	2	S1	C	NaN
680	10027572	R3	1	ZipLip Com Shop	10027572	G2	2		816-455-8733	3	S1	C	NaN
681	10027575	R3	1	Zitel Shop	10027575	G2	2		816-455-8733	2	S1	C	NaN
682	10027583	R2	1	zNET Shop	10027583	G2	2		816-455-8733	4	S5	C	NaN

The dataset consists of 3 excel sheets .

Description of data:

File1: Sales Data ------ it contains sales of all the items and sales amount
File2: Customer Address----- it contains address of the customers
File3: Customers------ it contains information of the customers.

```
In [14]: sales.columns
```

```
Out[14]: Index(['CustKey', 'DateKey', 'Discount Amount', 'Invoice Date',
       'Invoice Number', 'Item Class', 'Item Number', 'Item', 'Line Number',
       'List Price', 'Order Number', 'Promised Delivery Date', 'Sales Amount',
       'Sales Amount Based on List Price', 'Sales Cost Amount',
       'Sales Margin Amount', 'Sales Price', 'Sales Quantity', 'Sales Rep',
       'U/M', 'Unnamed: 20', 'Unnamed: 21'],
      dtype='object')
```

```
In [16]: customer_address.columns
```

```
Out[16]: Index(['Address Number', 'City', 'Country', 'Customer Address 1',
       'Customer Address 2', 'Customer Address 3', 'Customer Address 4',
       'State', 'Zip Code'],
      dtype='object')
```

```
In [15]: customer.columns
```

```
Out[15]: Index(['Address Number', 'Business Family', 'Business Unit', 'Customer',
       'Customer Number', 'Customer Type', 'Division', 'Line of Business',
       'Phone', 'Region Code', 'Regional Sales Mgr', 'Search Type',
       'Unnamed: 12'],
      dtype='object')
```

```
In [ ]:
```

```
In [ ]:
```

Here ,

1) Cust-key is primary key and address number is foreign key.

2) Date key = invoice date = promised delivery date .
3) Discount Amount = Sales Amount Based on List Price – Sales Amount
4) item ---- names of items
5) List Price --- sum of price of all items in list based on its MRP
6) Sales Amount---- it is the price given by the customer
7) Sales Amount Based on List Price= List Price * Sales Quantity
8) Sales Cost Amount---- it is an actual price paid by amazon to buy particular item.
9) Sales Margin Amount----Sales Amount- Sales Cost Amount
10) Sales Price ----- price of one item
11) Sales Quantity---- number of items bought

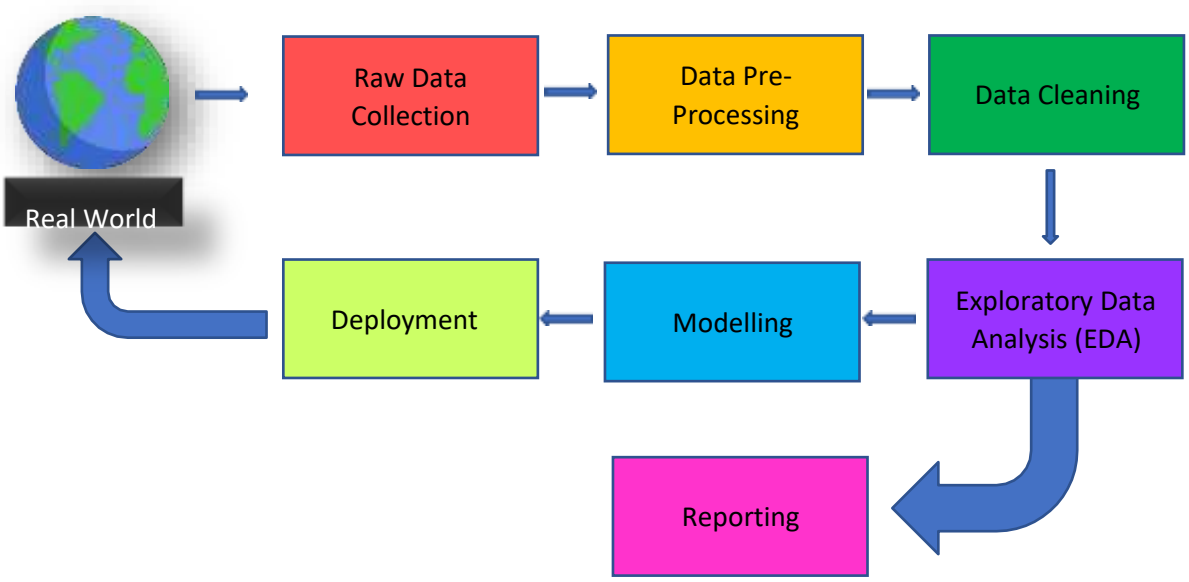
2. Problem Statement:

Investment is a game of understanding historic data of investment objects under different events but it is still a game of chances to minimize the risk we apply analytics to find the equilibrium investment.

This dataset contains year-wise data of sales

Year-wise sales analysis
quarter-wise sales analysis
monthly-wise sales analysis
Find key metrics and factors and show the meaningful relationships between attributes.
Do your own research and come up with your findings

3. Architecture



4.1 Architecture Description

1. **Raw Data Collection-** The Dataset was taken from iNeuron

2. **Data Pre-Processing**
   Before building any model, it is crucial to perform data pre-processing to feed the correct data to the model to learn and predict. Model performance depends on the quality of data to the model to train.
   This Process includes-
   a) Handling Null/ Missing Values

3. **Data Cleaning**
   Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
   a) Remove duplicate or irrelevant observations
   b) Filter unwanted outliers
   c) Renaming required attributes

4. **Exploratory Data Analysis (EDA)**
   Exploratory Data Analysis refers to the critical process of performing initial investigations on data to discover patterns, spot anomalies, test hypothesis

and to check assumptions with the help of summary statistics and graphical representations.

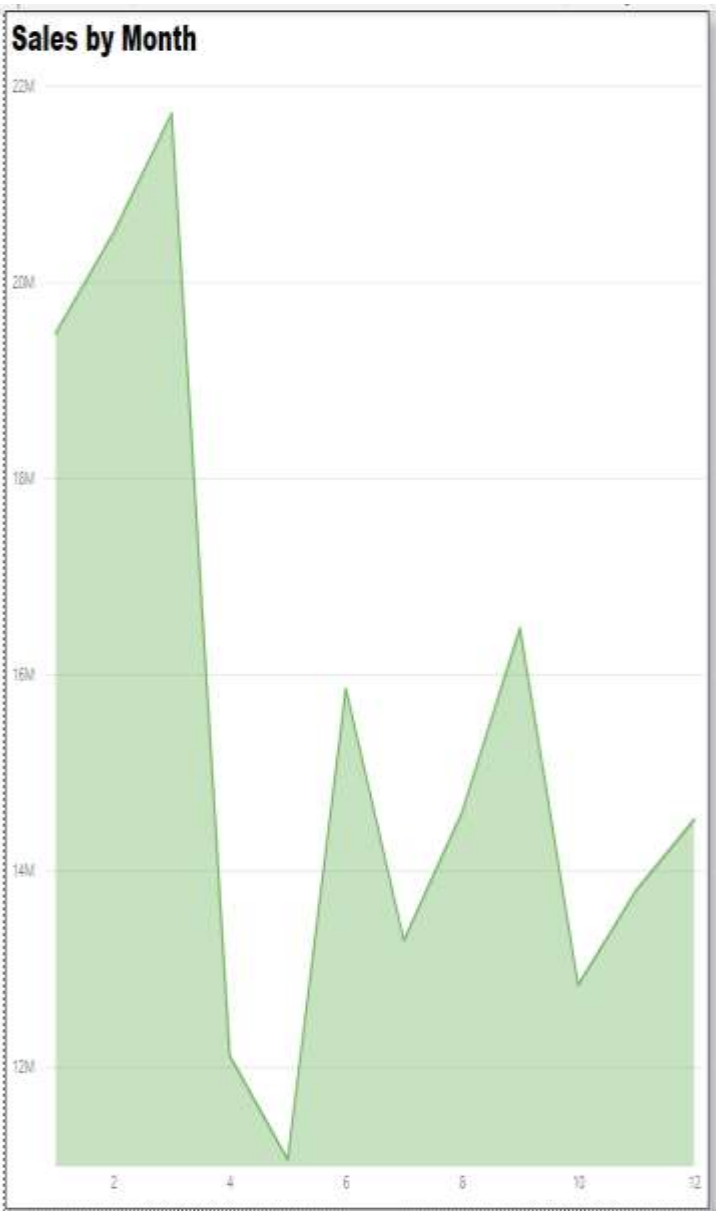
5. Reporting

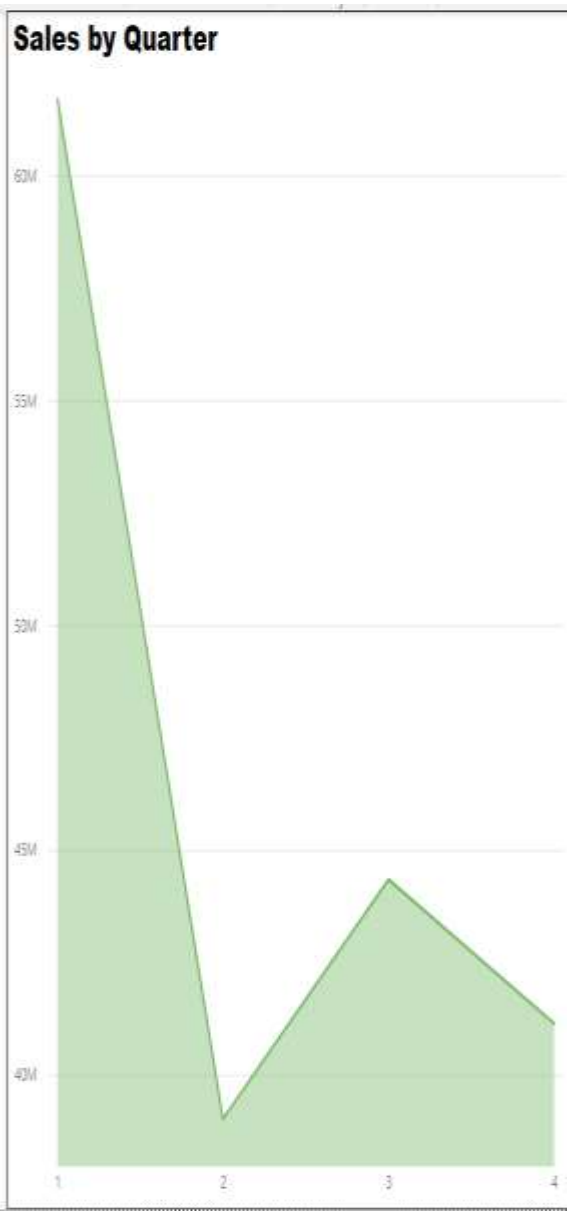
Reporting is a most important and underrated skill of a data analytics field. Because being a Data Analyst you should be good in easy in report because your model will be used by many stakeholders who are not from technical background.

a) High Level Design Document(HLD)
b) Low Level Design Document(LLD)
c) Architecture
d) Wireframe
e) Detailed Project Report
f) Power Point Presentation

6. Deployment

I have created Power Bi report :


Sales by Month
22M
20M
18M
16M
14M
12M
2
4
6
8
10
12


Sales by Quarter
60M
55M
50M
45M
40M
1
2
3
4