LOW LEVEL DESIGN DOCUMENT

(BANK MARKETING ANALYTICS – BI PROJECT)

TANUJ SHARMA

VERSION: 1.0
DATED: 28/12/2021

Document Version Control:

Bank Marketing Analytics - Business Intelligence Project

Version	Date	Author	Change
1.0	28/12/2021	Tanuj Sharma	First version of complete LLD
1.1	05/01/2022	Tanuj Sharma	Added Power BI functionality and dashboard

Abstract:

The data is related to direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe to a term deposit. The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be subscribed or not.

Contents:

1. Introduction:

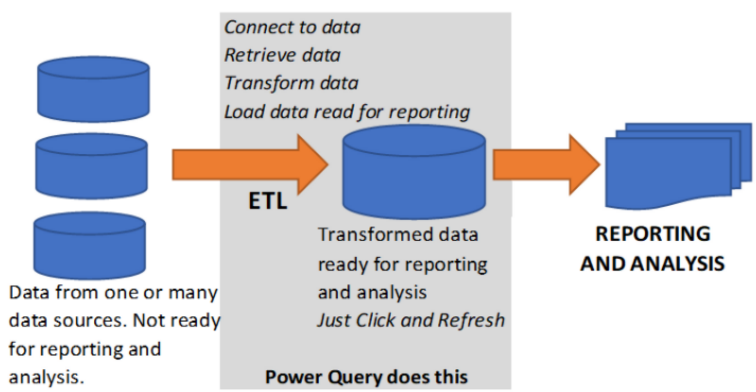
1.1. Why this Low-Level Design Document?

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 Bank Marketing Campaign Analysis. 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:



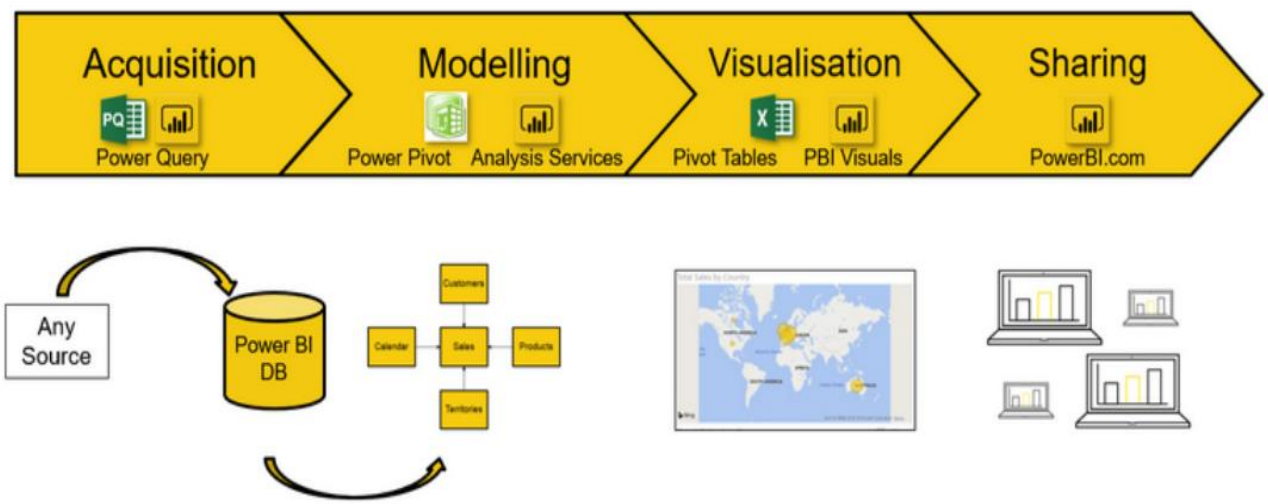
ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data.It also involves data visualization to draw meaningful patterns and insights.

Based on the results of ETL, companies also make business decisions, which can have repercussions later.

- If ETL is not done properly then it can damage the business a lot in many ways such as loss of client which we are working for, the decision making will go completely wrong and many more issues.
- If done well, it may improve the efficacy of everything we do next.

Below are following steps to follow for ETL:

1. Data Sourcing
2. Data Cleaning
3. Data Modelling
4. Data Visualization



3. Architecture Description:

3.1 Data Sourcing:

The dataset is in csv (comma separated values) format. MS Excel is used to load the data.

Citation Request:

This dataset is publicly available for research. The details are described in [Moro et al., 2014].
Please include this citation if you plan to use this database:

[Moro et al., 2014] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, In press, http://dx.doi.org/10.1016/j.dss.2014.03.001

Available at: [pdf] http://dx.doi.org/10.1016/j.dss.2014.03.001
[bib] http://www3.dsi.uminho.pt/pcortez/bib/2014-dss.txt

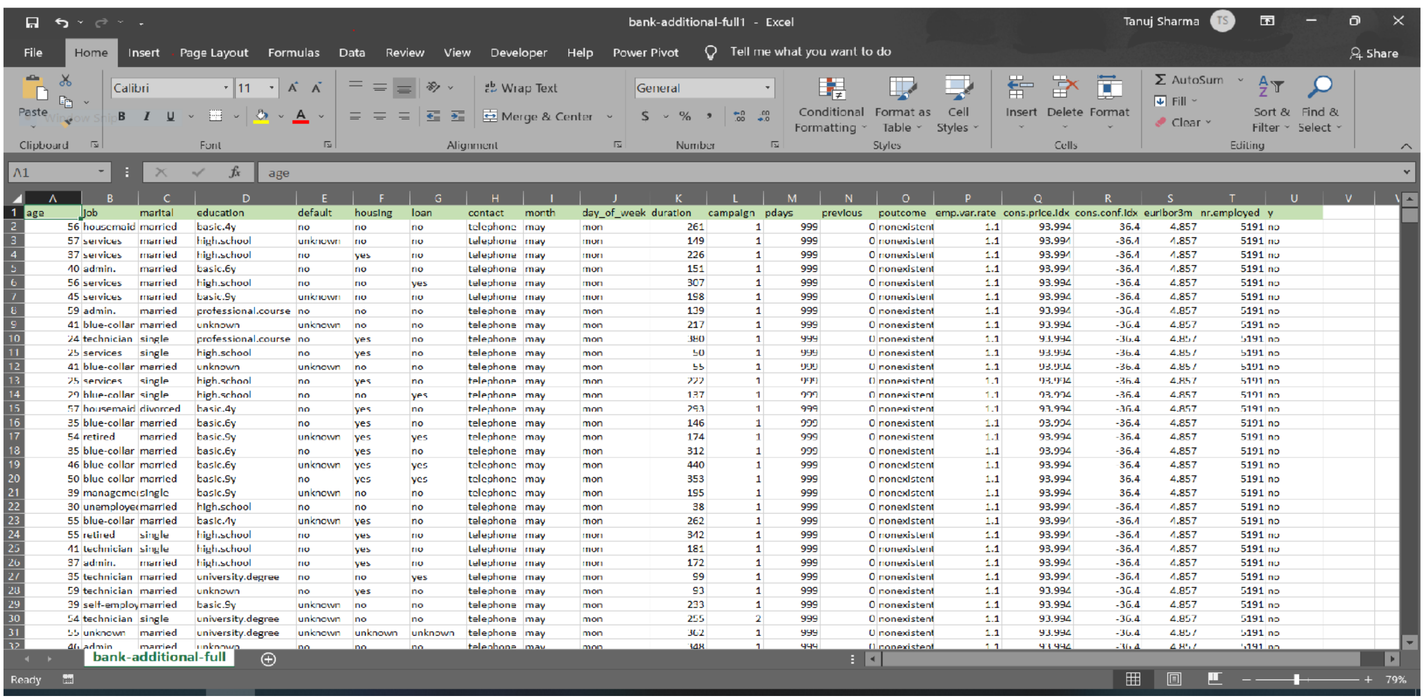
1. Title: Bank Marketing (with social/economic context)
2. Sources:
   Created by: Sérgio Moro (ISCTE-IUL), Paulo Cortez (Univ. Minho) and Paulo Rita (ISCTE-IUL) @ 2014
3. Past Usage:
   The full dataset (bank-additional-full.csv) was described and analyzed in:
   S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems (2014), doi:10.1016/j.dss.2014.03.001.

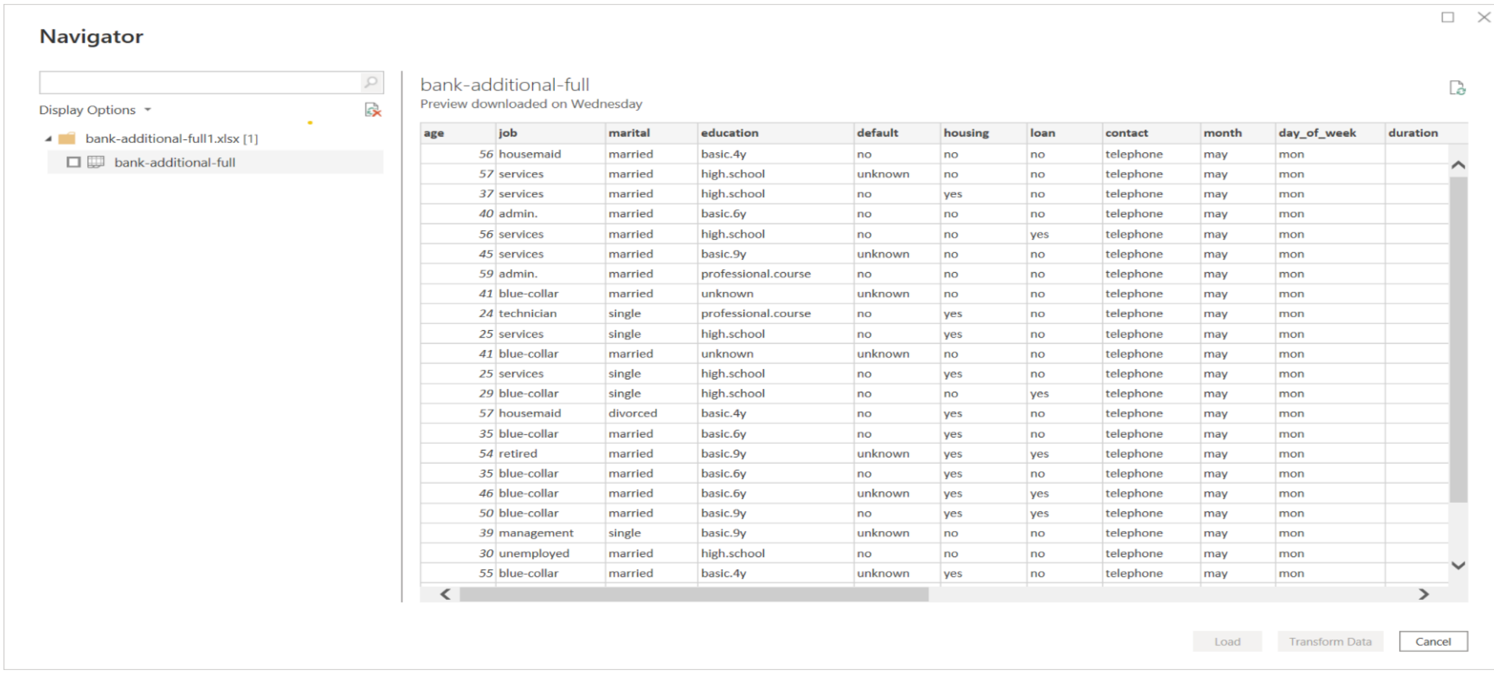
3.2 Data Overview:

- This dataset is based on "Bank Marketing" UCI dataset (please check the description at: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing).
- The data is enriched by the addition of five new social and economic features/attributes (national wide indicators from a ~10M population country), published by the Banco de Portugal and publicly available at: https://www.bportugal.pt/estatisticasweb.
- This dataset is almost identical to the one used in [Moro et al., 2014] (it does not include all attributes due to privacy concerns).
- Using the rminer package and R tool (http://cran.r-project.org/web/packages/rminer/), we found that the addition of the five new social and economic attributes (made available here) lead to substantial improvement in the prediction of a success, even when the duration of the call is not included. Note: the file can be read in R using: d=read.table("bank-additional-full.csv",header=TRUE,sep=";")

The zip file includes two datasets:

1) bank-additional-full.csv with all examples, ordered by date (from May 2008 to November 2010).
2) bank-additional.csv with 10% of the examples (4119), randomly selected from bank-additional-full.csv.
3) The smallest dataset is provided to test more computationally demanding machine learning algorithms (e.g., SVM).
4) The binary classification goal is to predict if the client will subscribe a bank term deposit (variable y).
5) Number of Instances: 41188 for bank-additional-full.csv
6) Number of Attributes: 20 + output attribute.





age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration
56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	
57	services	married	high.school	unknown	no	no	telephone	may	mon	
37	services	married	high.school	no	yes	no	telephone	may	mon	
40	admin.	married	basic.6y	no	no	no	telephone	may	mon	
56	services	married	high.school	no	no	yes	telephone	may	mon	
45	services	married	basic.9y	unknown	no	no	telephone	may	mon	
59	admin.	married	professional.course	no	no	no	telephone	may	mon	
41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	
24	technician	single	professional.course	no	yes	no	telephone	may	mon	
25	services	single	high.school	no	yes	no	telephone	may	mon	
41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	
25	services	single	high.school	no	yes	no	telephone	may	mon	
29	blue-collar	single	high.school	no	no	yes	telephone	may	mon	
57	housemaid	divorced	basic.4y	no	yes	no	telephone	may	mon	
35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	
54	retired	married	basic.9y	unknown	yes	yes	telephone	may	mon	
35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	
46	blue-collar	married	basic.6y	unknown	yes	yes	telephone	may	mon	
50	blue-collar	married	basic.9y	no	yes	yes	telephone	may	mon	
39	management	single	basic.9y	unknown	no	no	telephone	may	mon	
30	unemployed	married	high.school	no	no	no	telephone	may	mon	
55	blue-collar	married	basic.4y	unknown	yes	no	telephone	may	mon	

3.3 Data Description

Input variables:

Bank client data:

1 - age (numeric)
2 - job : type of job (categorical: "admin.","blue-collar","entrepreneur","housemaid","management","retired","self-employed","services","student","technician","unemployed","unknown")
3 - marital : marital status (categorical: "divorced","married","single","unknown"; note: "divorced" means divorced or widowed)
4 - education (categorical: "basic.4y","basic.6y","basic.9y","high.school","illiterate","professional.course","university.degree","unknown")
5 - default: has credit in default? (categorical: "no","yes","unknown")
6 - housing: has housing loan? (categorical: "no","yes","unknown")
7 - loan: has personal loan? (categorical: "no","yes","unknown")

related with the last contact of the current campaign:

8 - contact: contact communication type (categorical: "cellular","telephone")
9 - month: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")

10 - day_of_week: last contact day of the week (categorical: "mon","tue","wed","thu","fri")
11 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y="no"). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Other attributes:

12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
13 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
14 - previous: number of contacts performed before this campaign and for this client (numeric)
15 - poutcome: outcome of the previous marketing campaign (categorical: "failure","nonexistent","success")

Social and economic context attributes

16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)
17 - cons.price.idx: consumer price index - monthly indicator (numeric)
18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)
19 - euribor3m: euribor 3 month rate - daily indicator (numeric)
20 - nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):

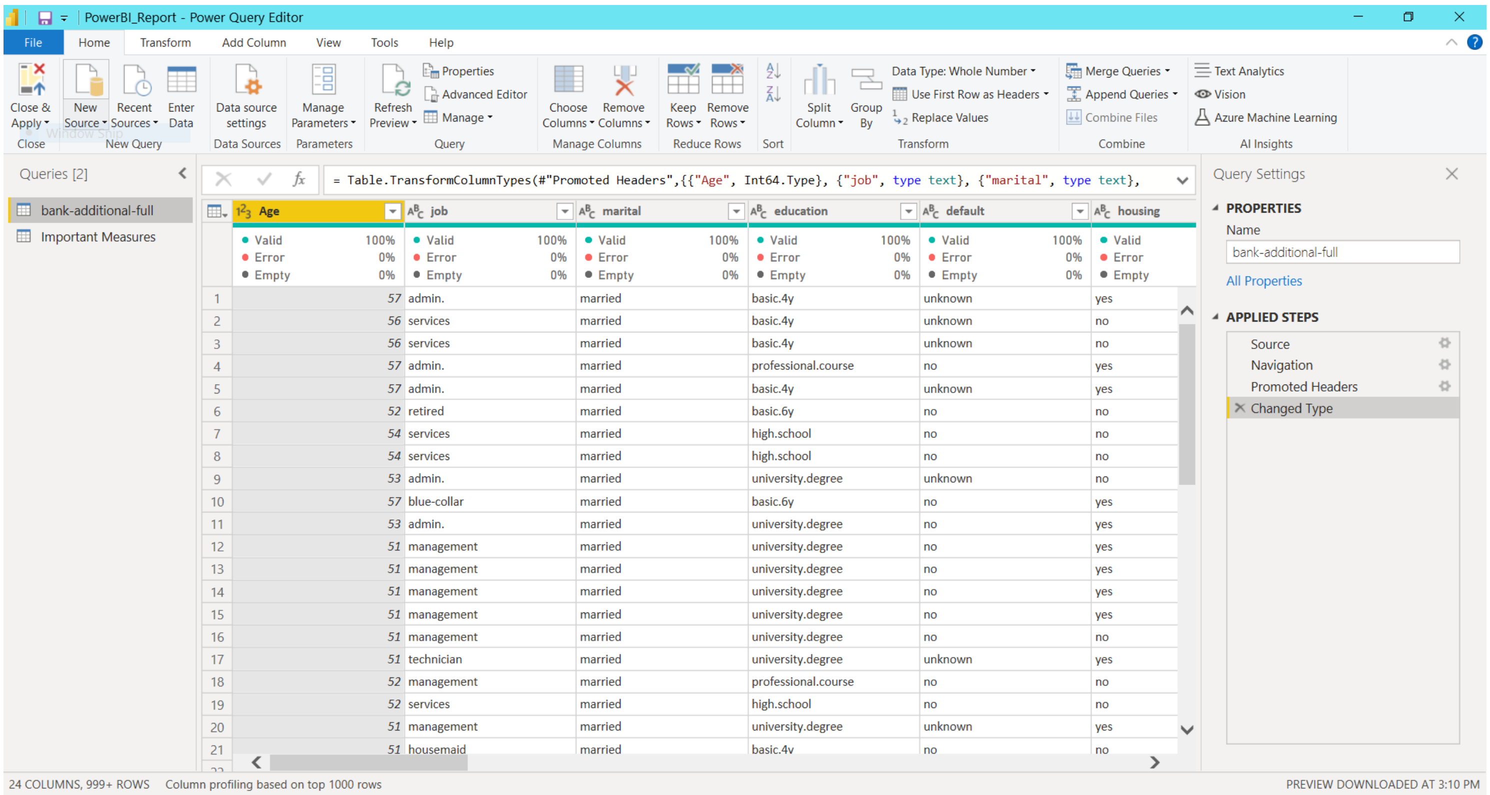
21 - y - has the client subscribed a term deposit? (binary: "yes","no")

3.4 Data loading in Power BI Query Editor

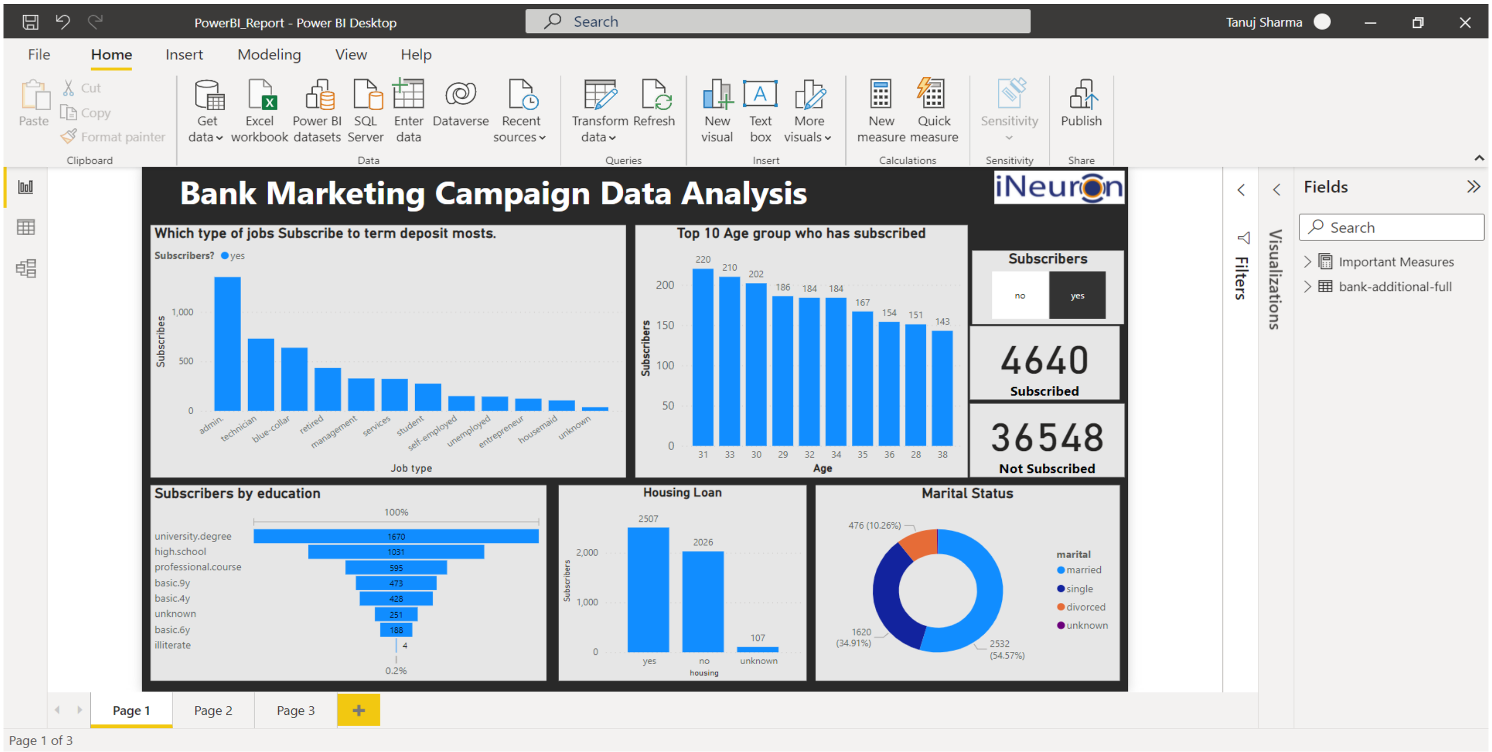
Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, dataverse, and more with the following characteristics:

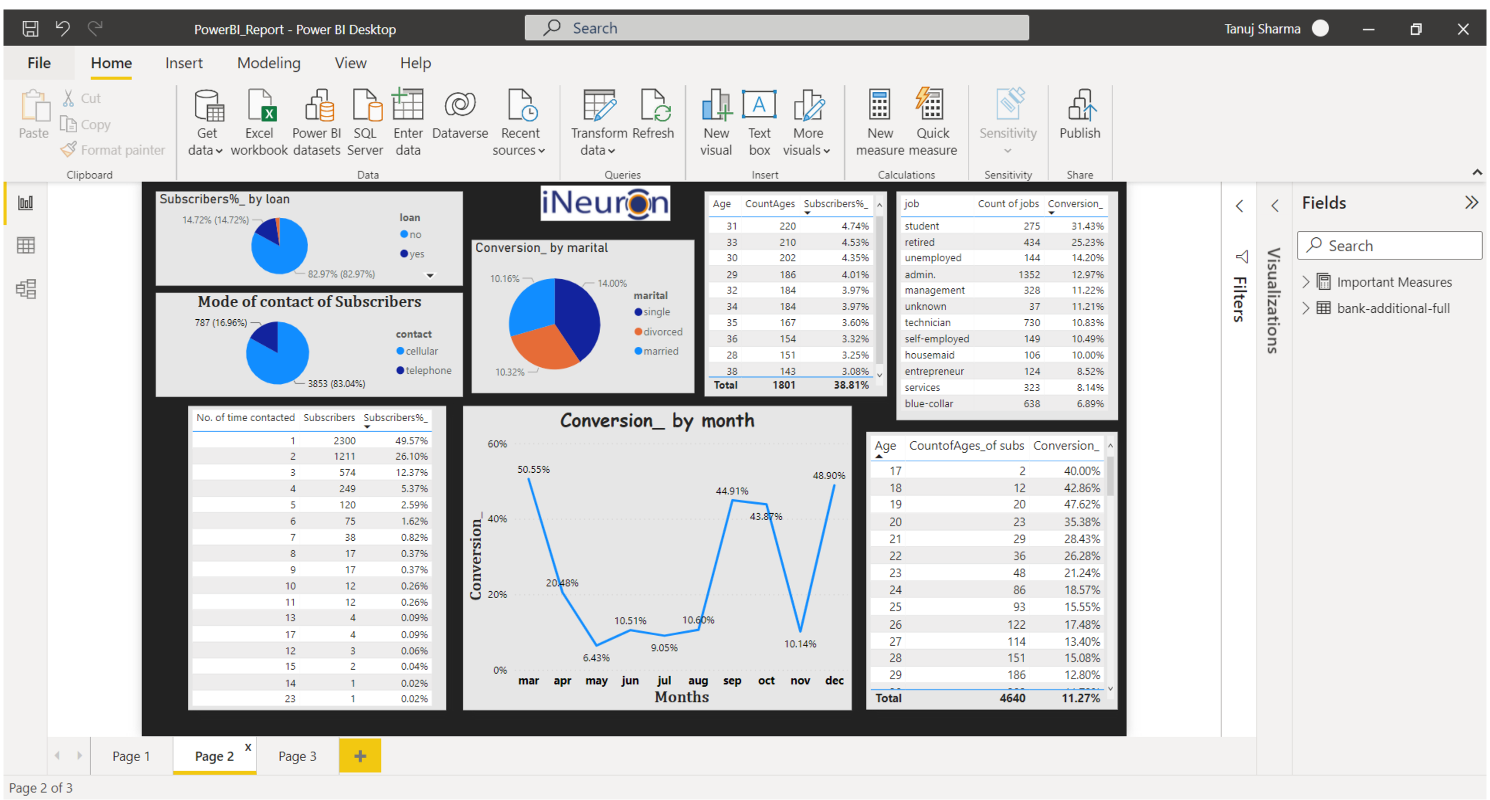
- There can be multiple rows and columns in the data.

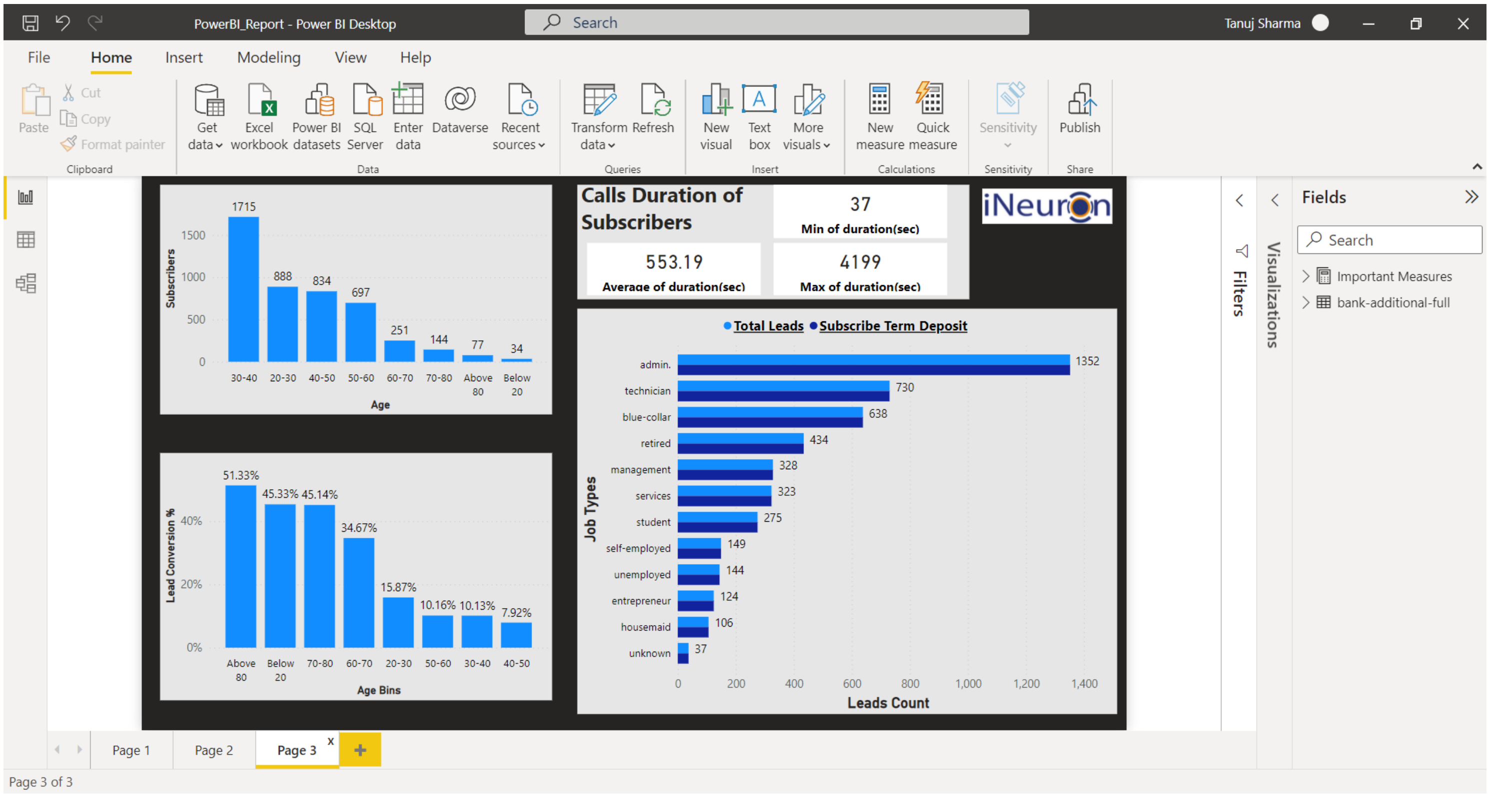
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a differents type of data - e.g. numbers, strings, dates, Boolean etc.



3.5 Data to Insights through Visualizations and Excel Data Analysis


PowerBI_Report - Power BI Desktop
Search
Tanuj Sharma
File
Home
Insert
Modeling
View
Help
Bank Marketing Campaign Data Analysis
iNeuron
Which type of jobs Subscribe to term deposit mosts.
Subscribers? yes
Subscribes
Job type
admin.
technician
blue-collar
retired
management
services
student
self-employed
unemployed
entrepreneur
housemaid
unknown
Top 10 Age group who has subscribed
Subscribers
Age
220 210 202 186 184 184 167 154 151 143
31 33 30 29 32 34 35 36 28 38
Subscribers
no
yes
4640
Subscribed
36548
Not Subscribed
Subscribers by education
100%
university.degree 1670
high.school 1031
professional.course 595
basic.9y 473
basic.4y 428
unknown 251
basic.6y 188
illiterate 4
0.2%
Housing Loan
2507
2026
107
yes
no
unknown
housing
Marital Status
476 (10.26%)
1620 (34.91%)
2532 (54.57%)
marital
married
single
divorced
unknown
Fields
Search
Important Measures
bank-additional-full
Filters
Visualizations
Page 1
Page 2
Page 3
Page 1 of 3


PowerBI_Report - Power BI Desktop
Search
Tanuj Sharma
Subscribers%_ by loan
14.72% (14.72%)
82.97% (82.97%)
loan
no
yes
iNeuron
Mode of contact of Subscribers
787 (16.96%)
3853 (83.04%)
contact
cellular
telephone
Conversion_ by marital
10.16%
14.00%
10.32%
marital
single
divorced
married
Age CountAges Subscribers%_
31 220 4.74%
33 210 4.53%
30 202 4.35%
29 186 4.01%
32 184 3.97%
34 184 3.97%
35 167 3.60%
36 154 3.32%
28 151 3.25%
38 143 3.08%
Total 1801 38.81%
job Count of jobs Conversion_
student 275 31.43%
retired 434 25.23%
unemployed 144 14.20%
admin. 1352 12.97%
management 328 11.22%
unknown 37 11.21%
technician 730 10.83%
self-employed 149 10.49%
housemaid 106 10.00%
entrepreneur 124 8.52%
services 323 8.14%
blue-collar 638 6.89%
No. of time contacted Subscribers Subscribers%_
1 2300 49.57%
2 1211 26.10%
3 574 12.37%
4 249 5.37%
5 120 2.59%
6 75 1.62%
7 38 0.82%
8 17 0.37%
9 17 0.37%
10 12 0.26%
11 12 0.26%
13 4 0.09%
17 4 0.09%
12 3 0.06%
15 2 0.04%
14 1 0.02%
23 1 0.02%
Conversion_ by month
Conversion
Months
50.55% 20.48% 6.43% 10.51% 9.05% 10.60% 44.91% 43.87% 10.14% 48.90%
mar apr may jun jul aug sep oct nov dec
Age CountofAges_of subs Conversion_
17 2 40.00%
18 12 42.86%
19 20 47.62%
20 23 35.38%
21 29 28.43%
22 36 26.28%
23 48 21.24%
24 86 18.57%
25 93 15.55%
26 122 17.48%
27 114 13.40%
28 151 15.08%
29 186 12.80%
Total 4640 11.27%
Page 1
Page 2
Page 3
Page 2 of 3



euribor3m	
Mean	3.6212908
Standard Error	0.0085463
Median	4.857
Mode	4.857
Standard Deviation	1.7344474
Sample Variance	3.0083078
Kurtosis	-1.4068026
Skewness	-0.709188
Range	4.411
Minimum	0.634
Maximum	5.045
Sum	149153.73
Count	41188

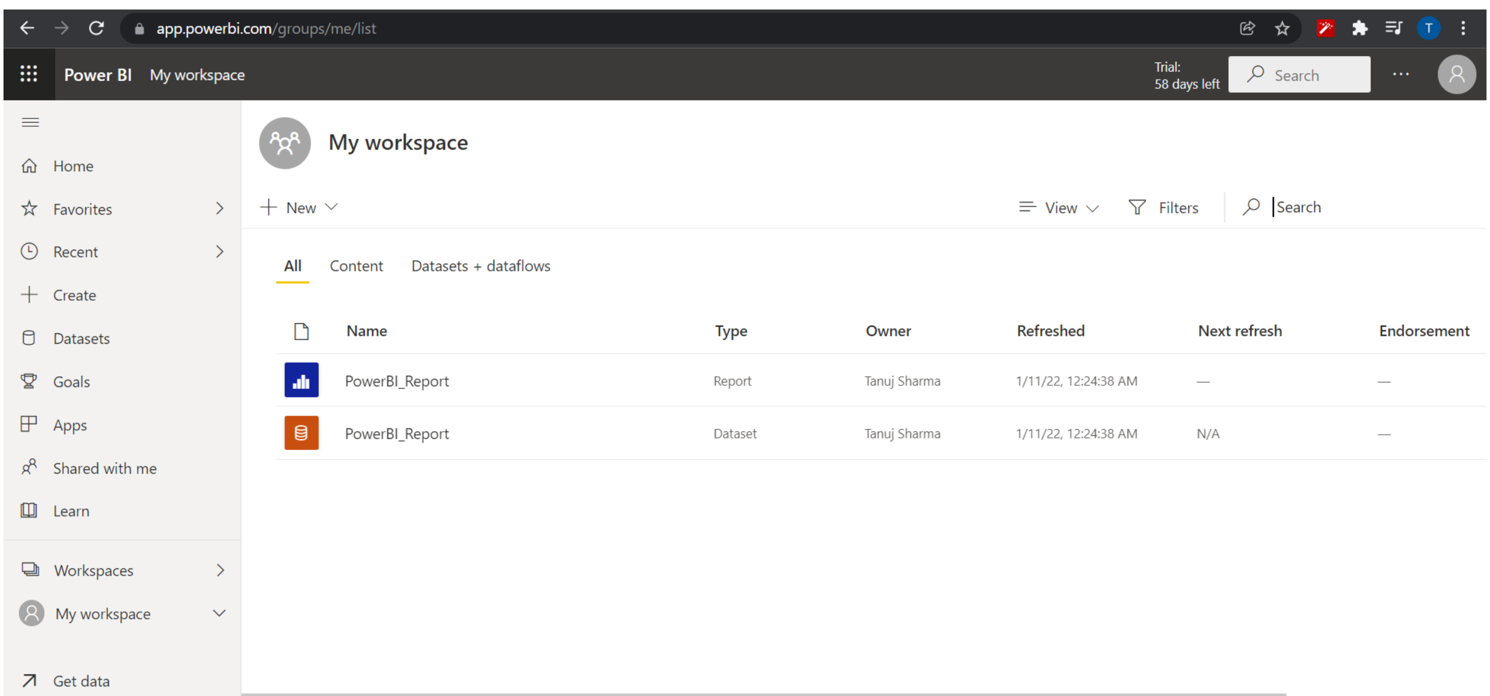
emp.var.rate	
Mean	0.0818855
Standard Error	0.0077407
Median	1.1
Mode	1.4
Standard Deviation	1.5709597
Sample Variance	2.4679145
Kurtosis	-1.0626315
Skewness	-0.7240955
Range	4.8
Minimum	-3.4
Maximum	1.4
Sum	3372.7
Count	41188

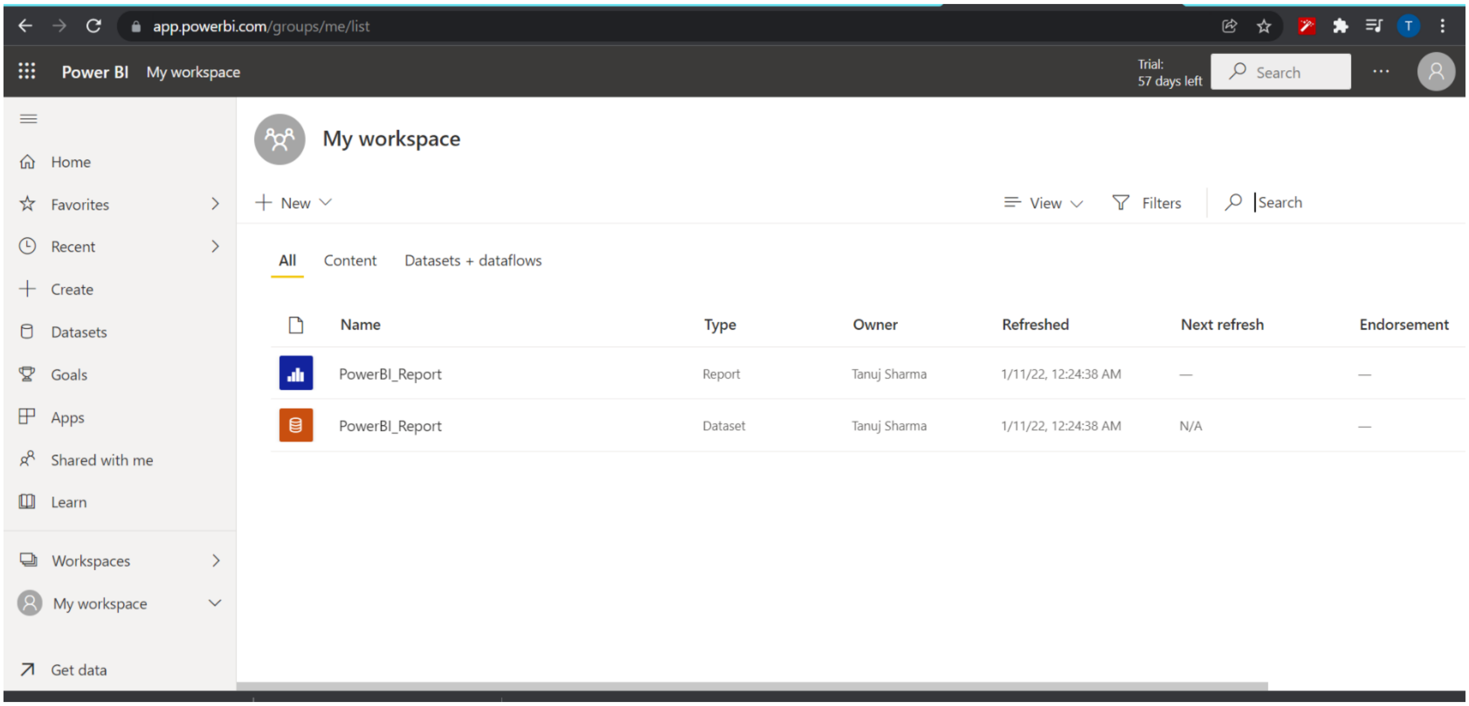
cons.price.idx	
Mean	93.575664
Standard Error	0.0028522
Median	93.749
Mode	93.994
Standard Deviation	0.57884
Sample Variance	0.3350558
Kurtosis	-0.8298086
Skewness	-0.2308877
Range	2.566
Minimum	92.201
Maximum	94.767
Sum	3854194.5
Count	41188

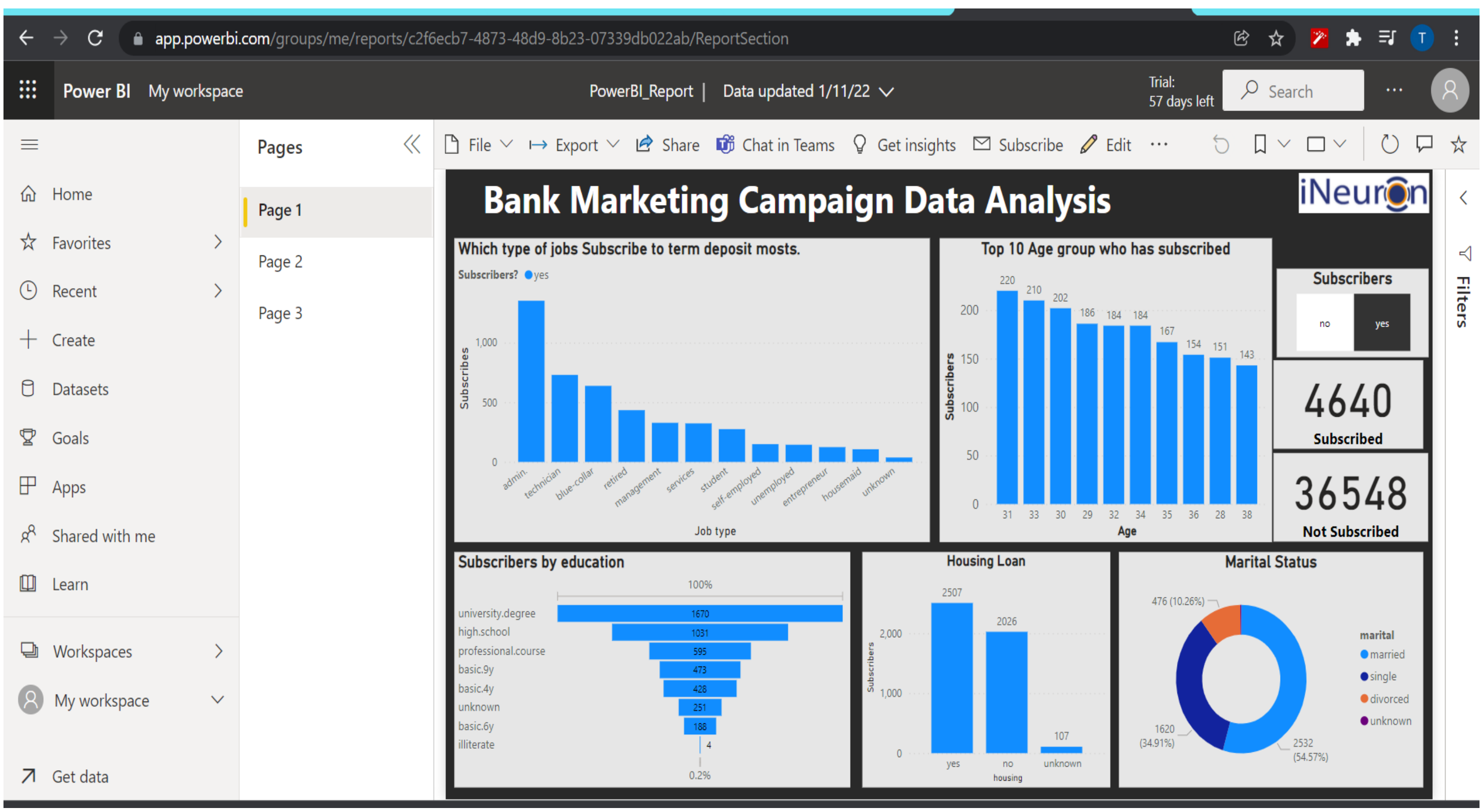
cons.conf.idx	
Mean	-40.5026
Standard Error	0.0228048
Median	-41.8
Mode	-36.4
Standard Deviation	4.6281979
Sample Variance	21.420215
Kurtosis	-0.3585583
Skewness	0.3031799
Range	23.9
Minimum	-50.8
Maximum	-26.9
Sum	-1668221.1
Count	41188

nr.employed	
Mean	5167.0359
Standard Error	0.3560096
Median	5191
Mode	5228.1
Standard Deviation	72.251528
Sample Variance	5220.2833
Kurtosis	-0.0037604
Skewness	-1.0442624
Range	264.5
Minimum	4963.6
Maximum	5228.1
Sum	212819875
Count	41188

4. Deployment to Power BI Service




app.powerbi.com/groups/me/list
Power BI
My workspace
Trial: 57 days left
Search
Home
Favorites
Recent
Create
Datasets
Goals
Apps
Shared with me
Learn
Workspaces
My workspace
Get data
My workspace
New
View
Filters
Search
All
Content
Datasets + dataflows
Name
Type
Owner
Refreshed
Next refresh
Endorsement
PowerBI_Report
Report
Tanuj Sharma
1/11/22, 12:24:38 AM
PowerBI_Report
Dataset
Tanuj Sharma
1/11/22, 12:24:38 AM
N/A


app.powerbi.com/groups/me/reports/c2f6ecb7-4873-48d9-8b23-07339db022ab/ReportSection
Power BI
My workspace
PowerBI_Report | Data updated 1/11/22
Trial: 57 days left
Search
Pages
Page 1
Page 2
Page 3
File
Export
Share
Chat in Teams
Get insights
Subscribe
Edit
Bank Marketing Campaign Data Analysis
iNeuron
Which type of jobs Subscribe to term deposit mosts.
Job type
Top 10 Age group who has subscribed
Age
Subscribers
4640
Subscribed
36548
Not Subscribed
Subscribers by education
Housing Loan
Marital Status
Filters