

ARCHITECTURE DESIGN DOCUMENT

(BANK MARKETING ANALYTICS – BI PROJECT)

TANUJ SHARMA

VERSION: 1.0
DATED: 09/01/2022



Document Version Control:

Bank Marketing Analytics - Business Intelligence Project

Version	Date	Author	Change
1.0	09/01/2022	Tanuj Sharma	First version of complete Architecture Design Document

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.....	05
1.1. What is Architecture Design Document?.....	05
1.2. Scope.....	05
2. Architecture.....	05
3. Architecture Description.....	06
3.1. Data Sourcing.....	06
3.2. Data Overview.....	07-08
3.3. Data Description.....	09
3.4. Data loading in Power BI Query Editor.....	10
3.5. Data to Insights through Visualizations and Excel Data Analysis.....	11-13
4. Deployment	14

1. Introduction:

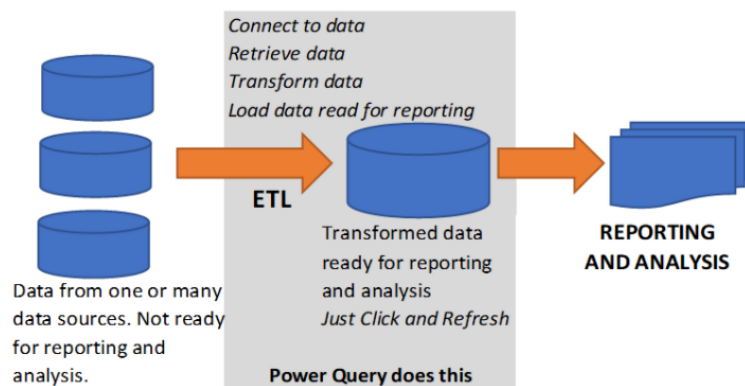
1.1. Why this Low-Level Design Document?

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:



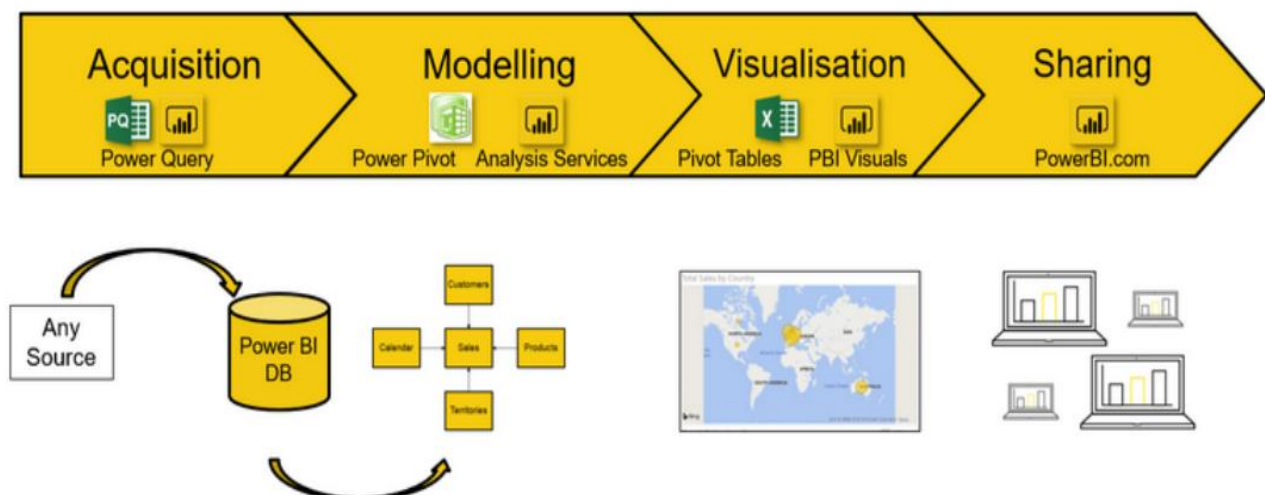
ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
	age	job	marital	education	default	housing	loan	contact	month	day_of_week	duration	campaign	pdays	previous	outcome	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y	
1	56	housemaid	married	basic.4y	no	no	no	telephone	may	mon	261	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
2	57	services	married	highschool	unknown	no	no	telephone	may	mon	149	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
3	37	services	married	highschool	no	yes	no	telephone	may	mon	226	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
4	40	admin.	married	basic.6y	no	no	no	telephone	may	mon	151	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
5	56	services	married	highschool	no	no	yes	telephone	may	mon	307	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
6	45	services	married	basic.5y	unknown	no	no	telephone	may	mon	158	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
7	59	admin.	married	professional.course	no	no	no	telephone	may	mon	139	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
8	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	217	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
9	24	technician	single	professional.course	no	yes	no	telephone	may	mon	380	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
10	25	services	single	highschool	no	yes	no	telephone	may	mon	50	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
11	41	blue-collar	married	unknown	unknown	no	no	telephone	may	mon	55	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
12	24	services	single	highschool	no	yes	no	telephone	may	mon	227	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
13	29	blue-collar	single	highschool	no	no	yes	telephone	may	mon	137	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
14	57	housemaid	divorced	basic.4y	no	yes	no	telephone	may	mon	293	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
15	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	146	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
16	54	retired	married	basic.5y	unknown	yes	yes	telephone	may	mon	174	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
17	35	blue-collar	married	basic.6y	no	yes	no	telephone	may	mon	312	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
18	46	blue collar	married	basic.6y	unknown	yes	yes	telephone	may	mon	440	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
19	50	blue collar	married	basic.5y	no	yes	yes	telephone	may	mon	353	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
20	39	manager	single	basic.5y	unknown	no	no	telephone	may	mon	155	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
21	30	unemployed	married	highschool	no	no	no	telephone	may	mon	38	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
22	55	blue-collar	married	basic.4y	unknown	yes	no	telephone	may	mon	262	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
23	55	retired	single	highschool	no	yes	no	telephone	may	mon	342	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
24	41	technician	single	highschool	no	yes	no	telephone	may	mon	181	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
25	37	admin.	married	highschool	no	yes	no	telephone	may	mon	172	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
26	35	technician	married	university.degree	no	no	yes	telephone	may	mon	59	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
27	29	technician	married	unknown	no	yes	no	telephone	may	mon	53	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
28	39	self-employed	married	basic.5y	unknown	no	no	telephone	may	mon	233	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
29	54	technician	single	university.degree	unknown	no	no	telephone	may	mon	255	2	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
30	25	unknown	married	university.degree	unknown	unknown	unknown	telephone	may	mon	362	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	
31	40	admin	married	unknown	no	no	no	telephone	may	mon	148	1	999	0	nonexistent	1.1	93.994	-36.4	4.857	5191	no	

Navigator

bank-additional-full
Preview downloaded on Wednesday

Display Options ▾

bank-additional-full1.xlsx [1]
bank-additional-full

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	

Load Transform Data Cancel

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:

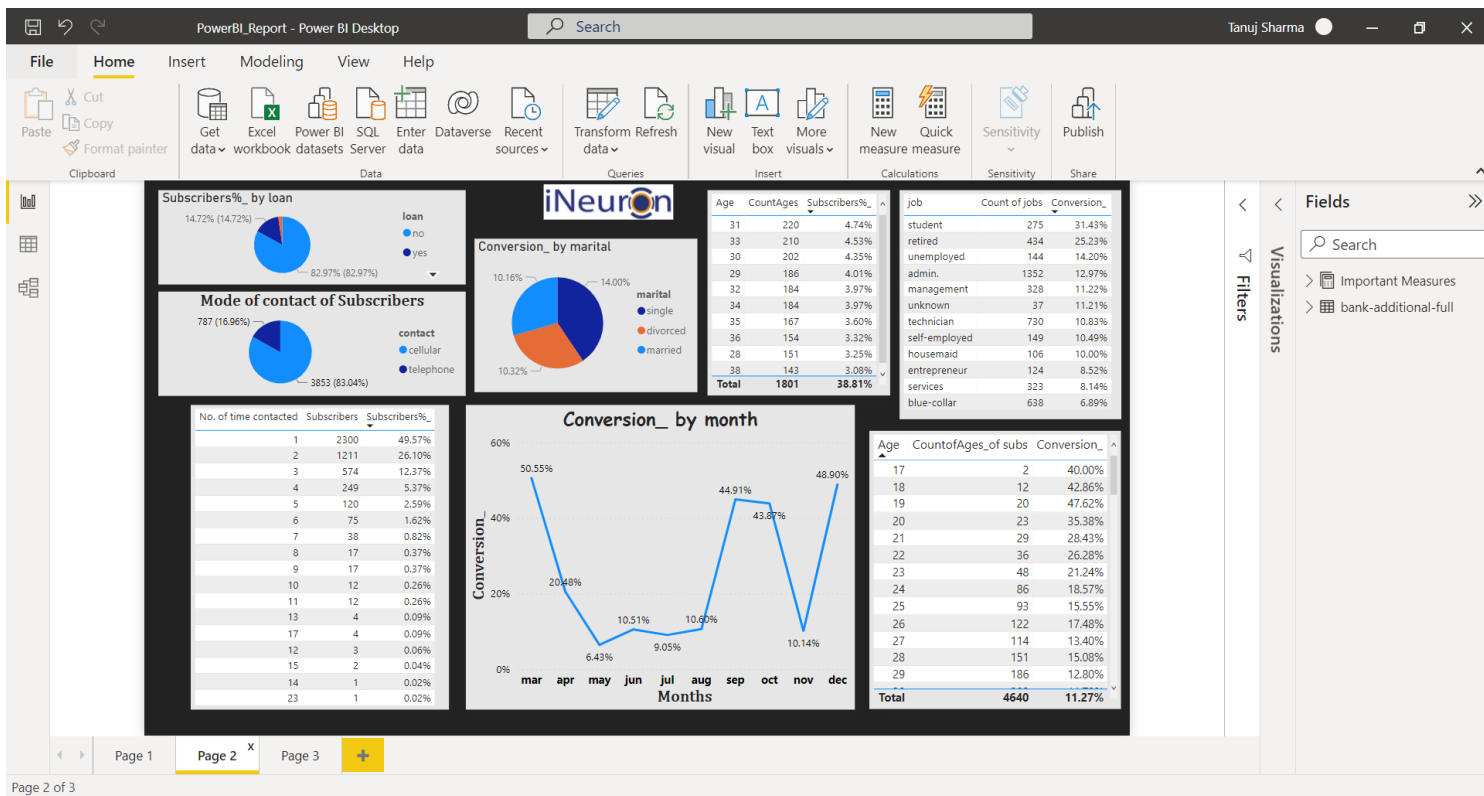
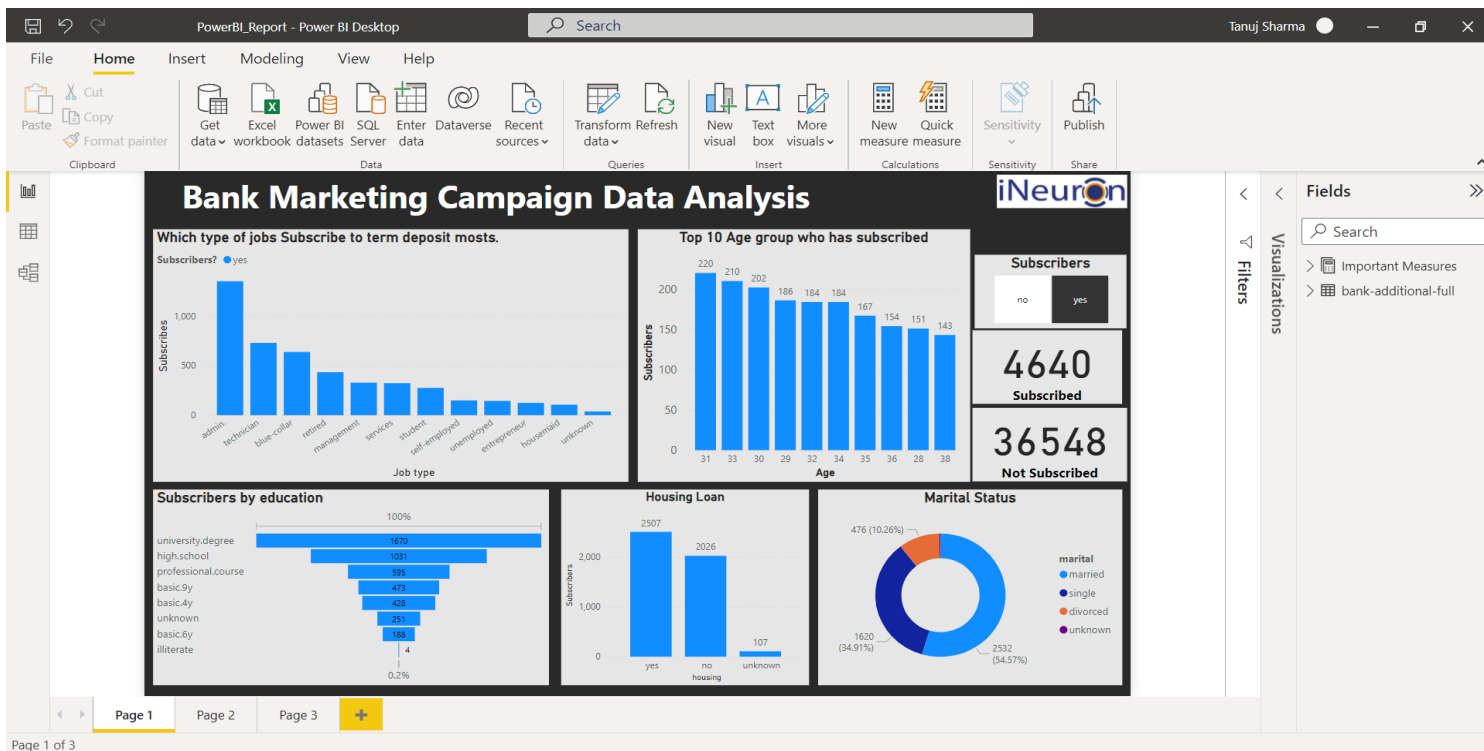
- There can be multiple rows and columns in the data.

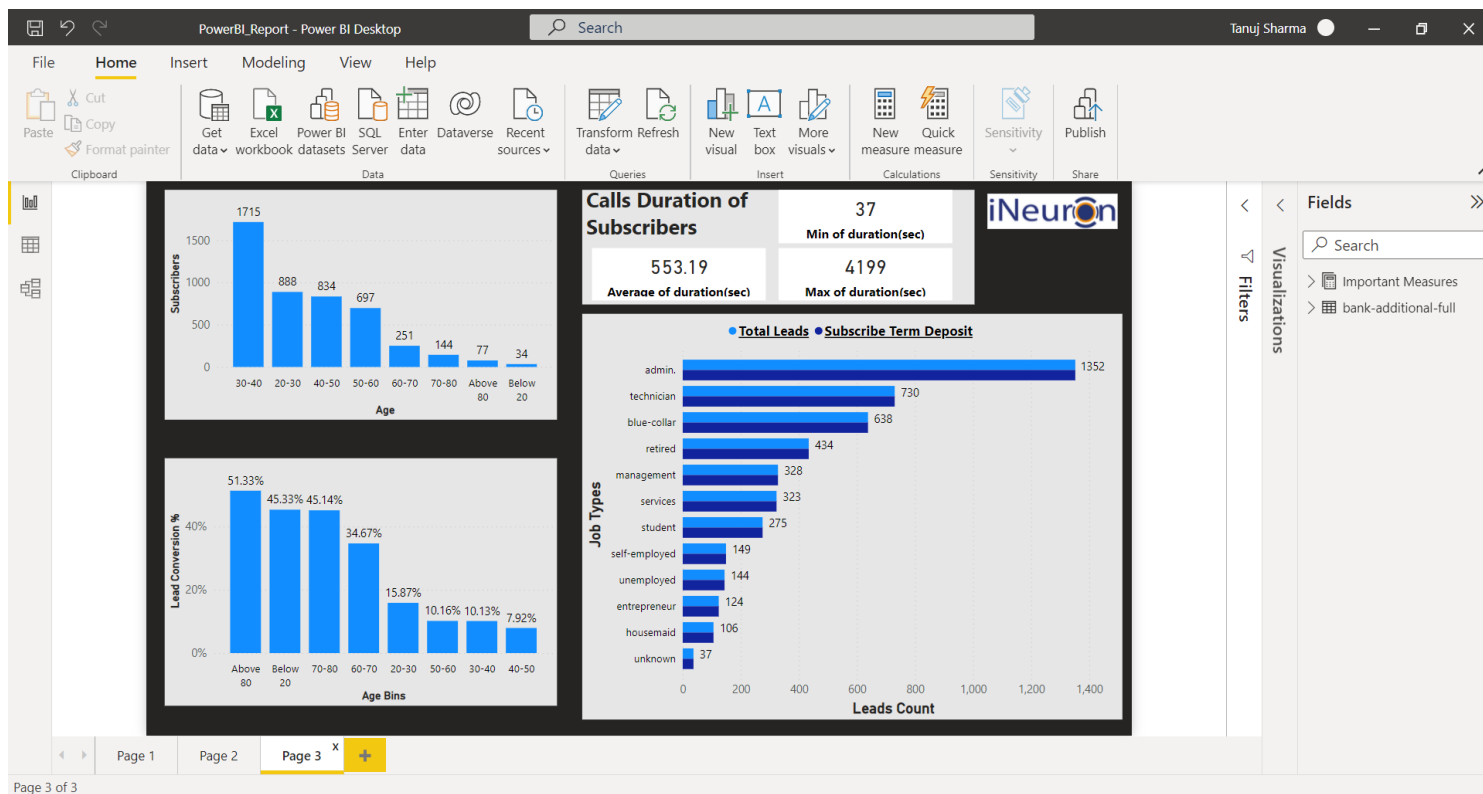
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data - e.g. numbers, strings, dates, Boolean etc.

The screenshot displays the Power Query Editor interface for a query named "bank-additional-full". The main area shows a table with 24 columns and 999+ rows. The columns are: Age, job, marital, education, default, and housing. Each column has a data type and a status bar indicating the percentage of valid, error, and empty data. The right-hand pane shows the "Query Settings" for the query, including the "Name" (bank-additional-full) and the "Applied Steps" (Source, Navigation, Promoted Headers, Changed Type).

	Age	job	marital	education	default	housing
1	57	admin.	married	basic.4y	unknown	yes
2	56	services	married	basic.4y	unknown	no
3	56	services	married	basic.4y	unknown	no
4	57	admin.	married	professional.course	no	yes
5	57	admin.	married	basic.4y	unknown	yes
6	52	retired	married	basic.6y	no	no
7	54	services	married	high.school	no	no
8	54	services	married	high.school	no	no
9	53	admin.	married	university.degree	unknown	no
10	57	blue-collar	married	basic.6y	no	yes
11	53	admin.	married	university.degree	no	yes
12	51	management	married	university.degree	no	yes
13	51	management	married	university.degree	no	yes
14	51	management	married	university.degree	no	yes
15	51	management	married	university.degree	no	yes
16	51	management	married	university.degree	no	no
17	51	technician	married	university.degree	unknown	yes
18	52	management	married	professional.course	no	no
19	52	services	married	high.school	no	no
20	51	management	married	university.degree	unknown	yes
21	51	housemaid	married	basic.4y	no	no

3.5 Data to Insights through Visualizations and Excel Data Analysis





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

The screenshot displays the Power BI Service interface for a workspace named 'My workspace'. The left sidebar contains navigation links: Home, Favorites, Recent, Create, Datasets, Goals, Apps, Shared with me, Learn, Workspaces, My workspace, and Get data. The main area shows a list of items in the workspace, with tabs for 'All', 'Content', and 'Datasets + dataflows'. The 'All' tab is selected, showing a table with the following data:

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/list

Power BI My workspace

Trial: 57 days left

Search

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

Which type of jobs Subscribe to term deposit mosts.

Subscribers? yes

Job type	Subscribers
admin	1000
technician	700
blue-collar	600
retired	400
management	300
services	250
student	200
self-employed	150
unemployed	100
entrepreneur	50
housemaid	20
unknown	10

Top 10 Age group who has subscribed

Age	Subscribers
31	220
33	210
30	202
29	186
32	184
34	184
35	167
36	154
28	151
38	143

Subscribers

no yes

4640 Subscribed

36548 Not Subscribed

Subscribers by education

Education	Subscribers
university.degree	1670
high.school	1031
professional.course	595
basic.9y	473
basic.4y	428
unknown	251
basic.6y	188
illiterate	4

Housing Loan

Housing	Subscribers
yes	2507
no	2026
unknown	107

Marital Status

Marital	Subscribers	Percentage
married	2532	54.57%
divorced	476	10.26%
single	1620	34.91%
unknown	10	0.22%