ARCHITECTURE DESIGN DOCUMENT

(BANK MARKETING ANALYTICS – BI PROJECT)

TAJ HASAN MANSURI

VERSION: 1.0
DATED: 22/01/2023

Document Version Control:

Bank Marketing Analytics - Business Intelligence Project

Version	Date	Author	Change
1.0	12/01/2023	Taj Hasan Mansuri	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:

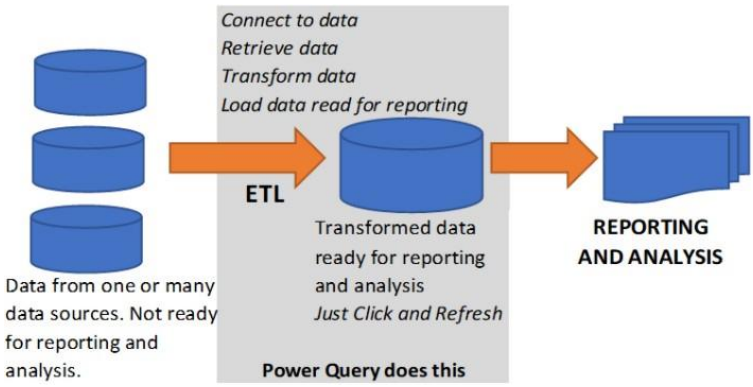
1.1. Why this Low-Level Design Document?

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:



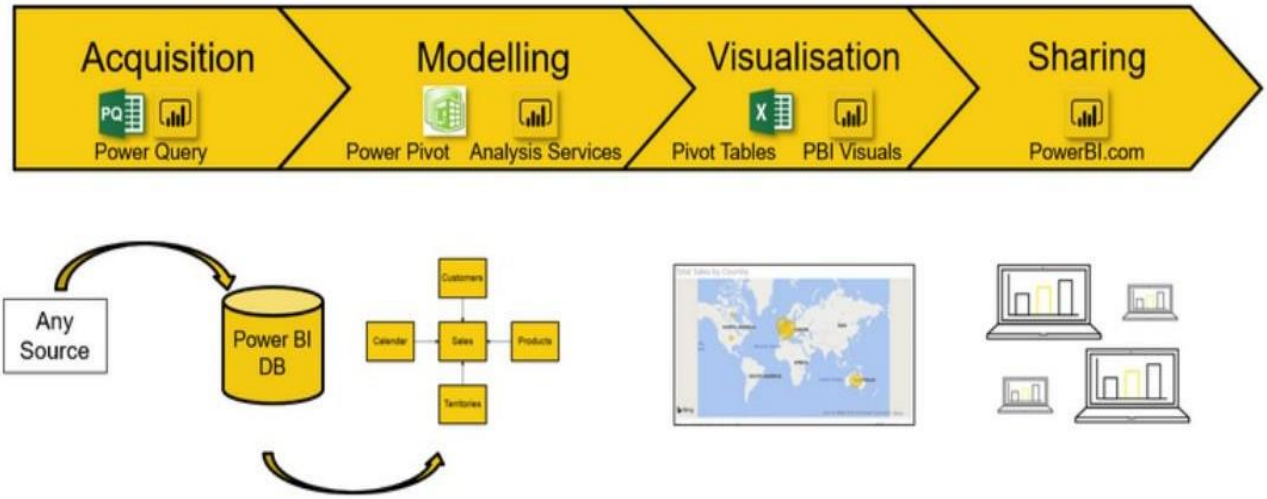
ETL (extract, transform and load) in Power BI uses preparation of data sets for analysis by removing irregularities in the data. It also involves data visualization to draw meaningful patterns and insights.

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. MySQL Workbench 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:

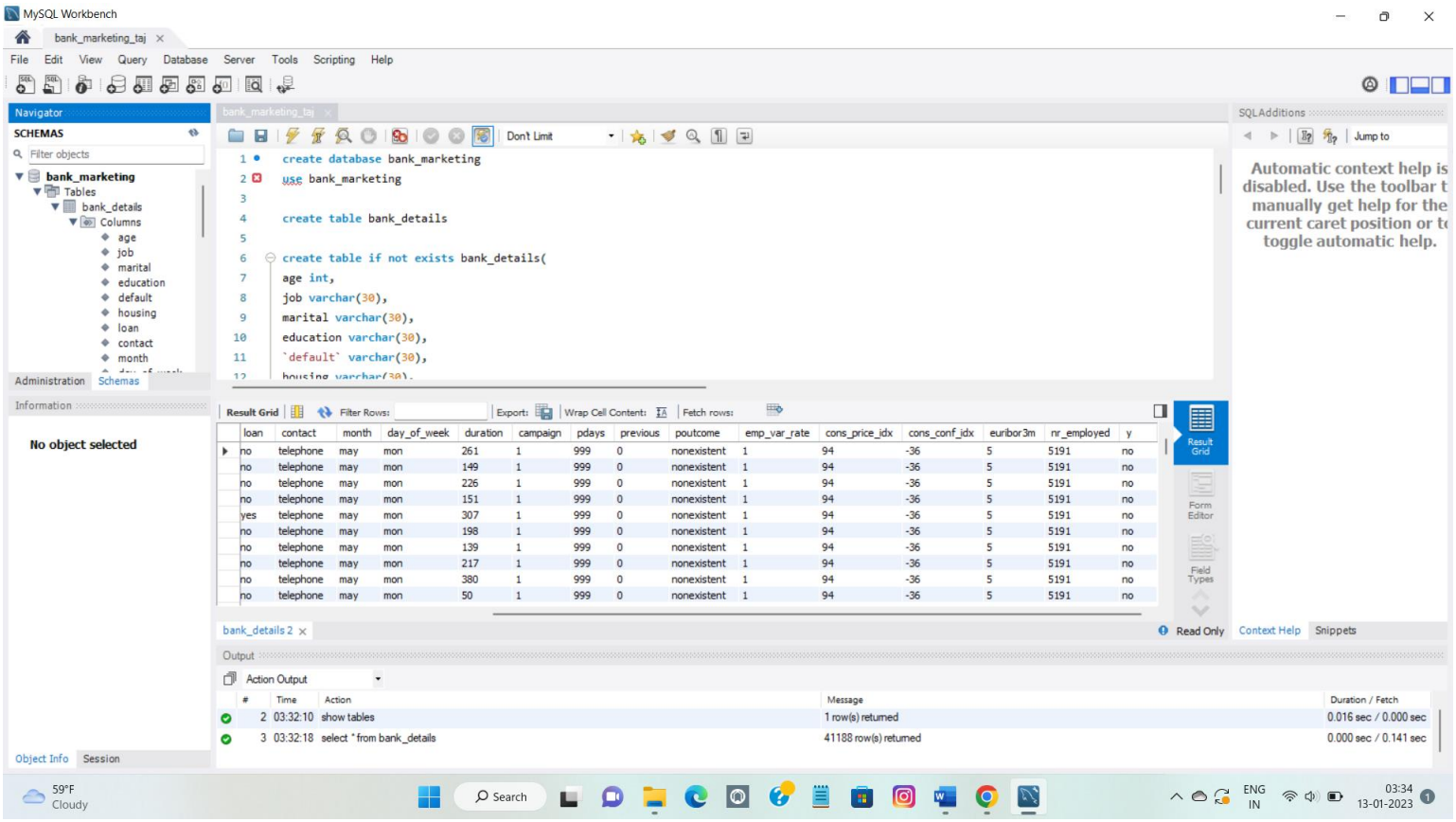
   S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems (2014) doi:10.1016/j.dss.2014.03.001.

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.



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)

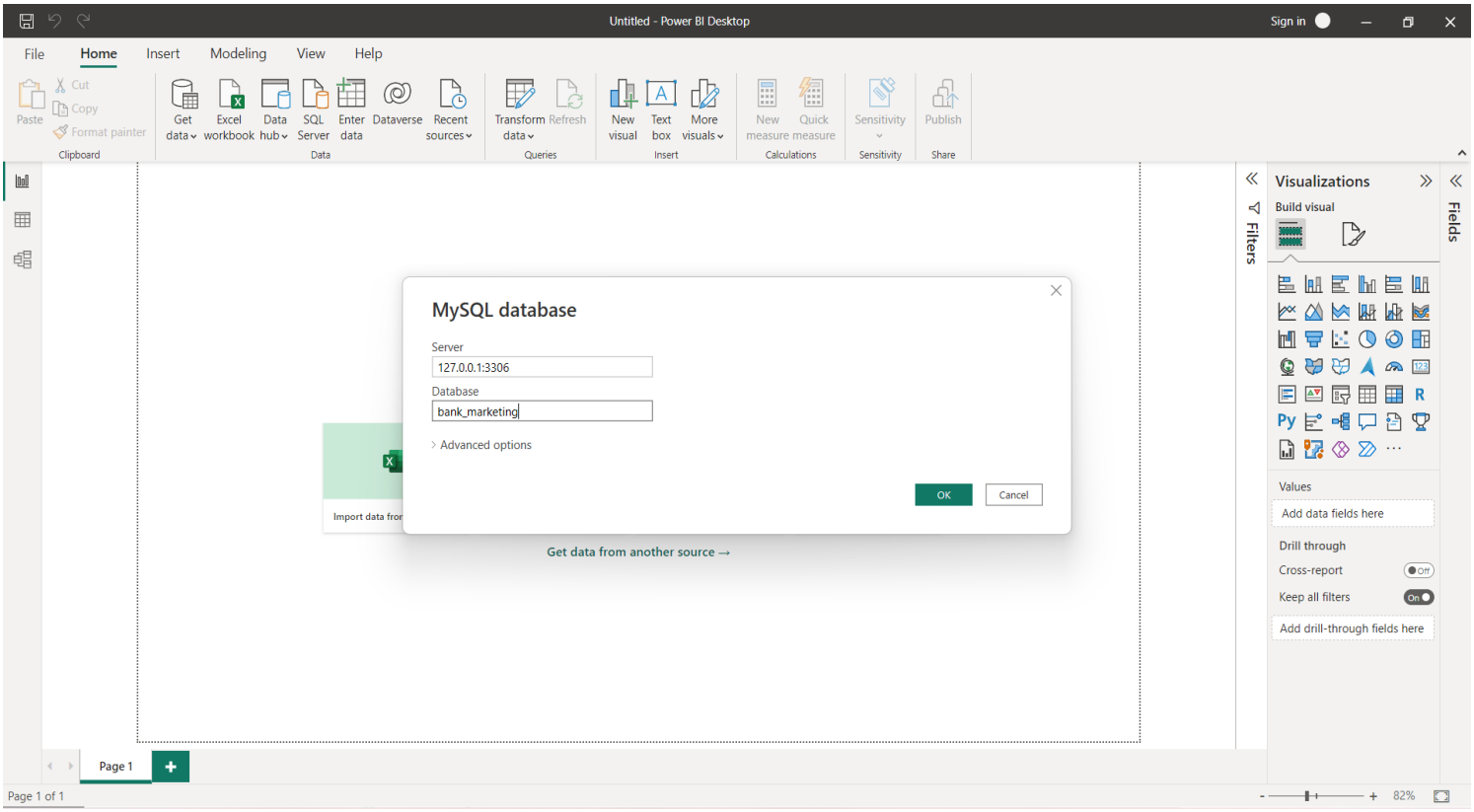
#Output variable (desired target):

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

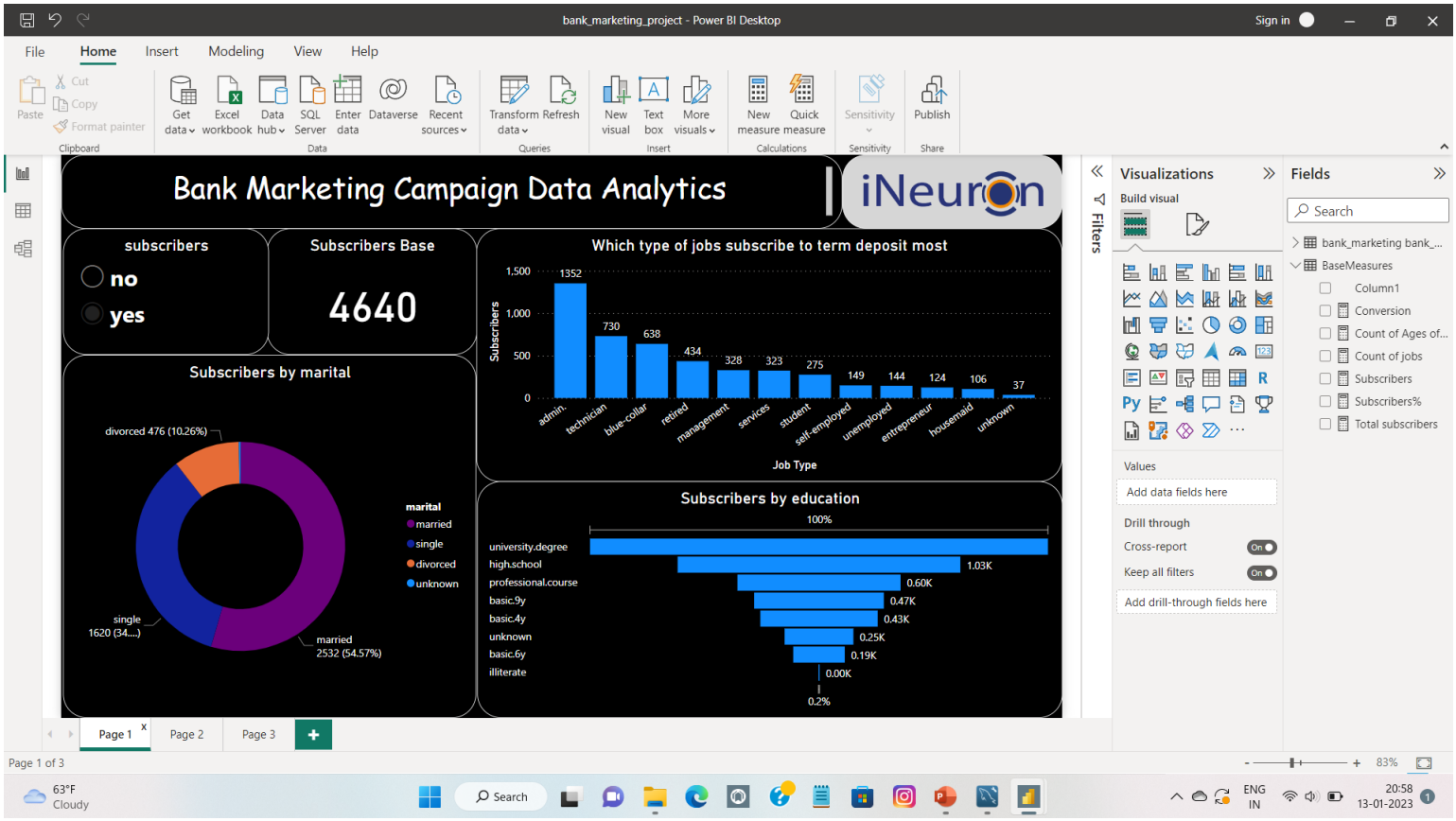
3.4 Data loading in Power BI Query Editor

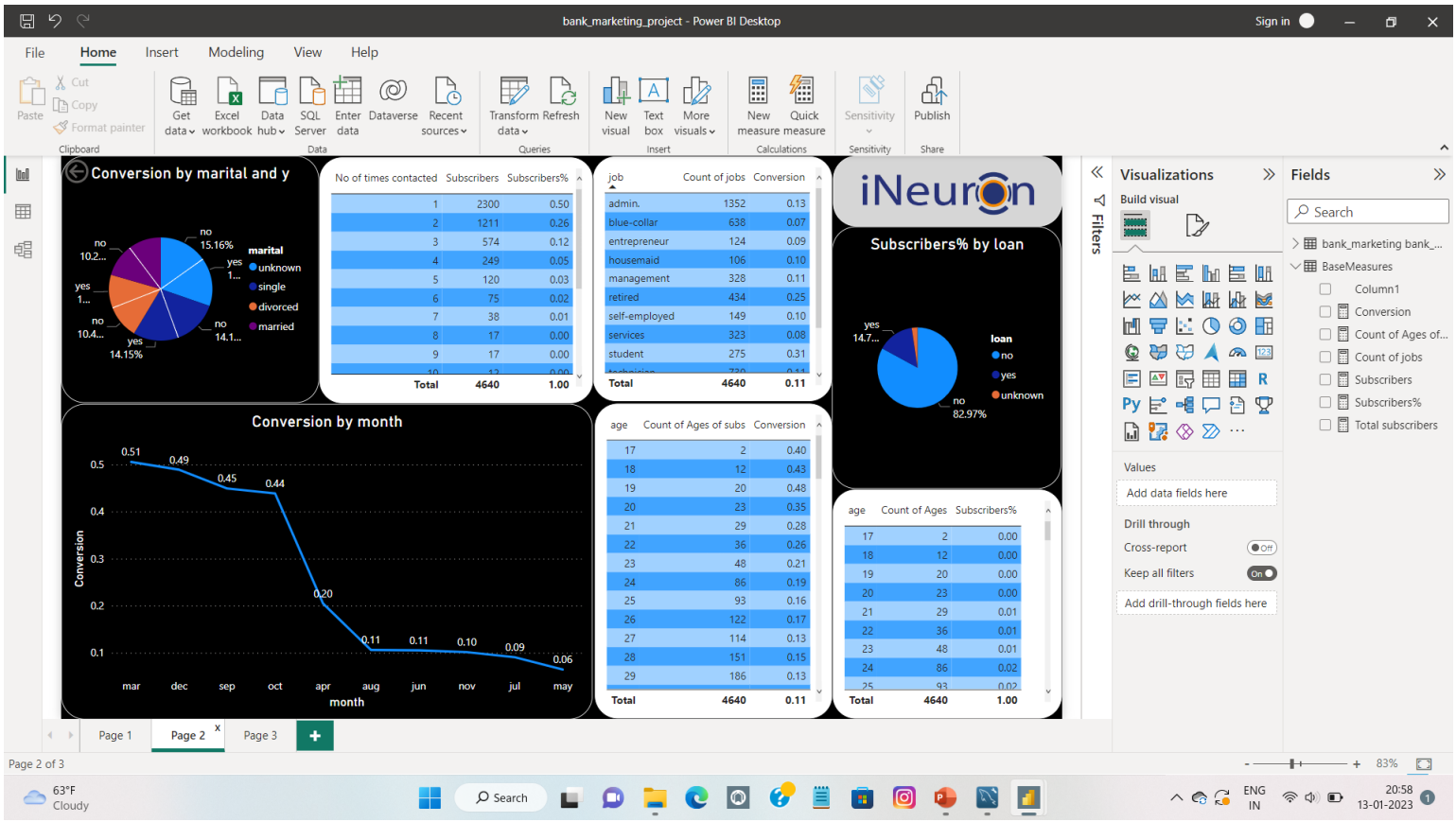
Power Query is the data connectivity and data preparation technology that enables end users to seamlessly import and reshape data from within a wide range of Microsoft products, including Excel, Power BI, Analysis Services, data verse, and more with the following characteristics:

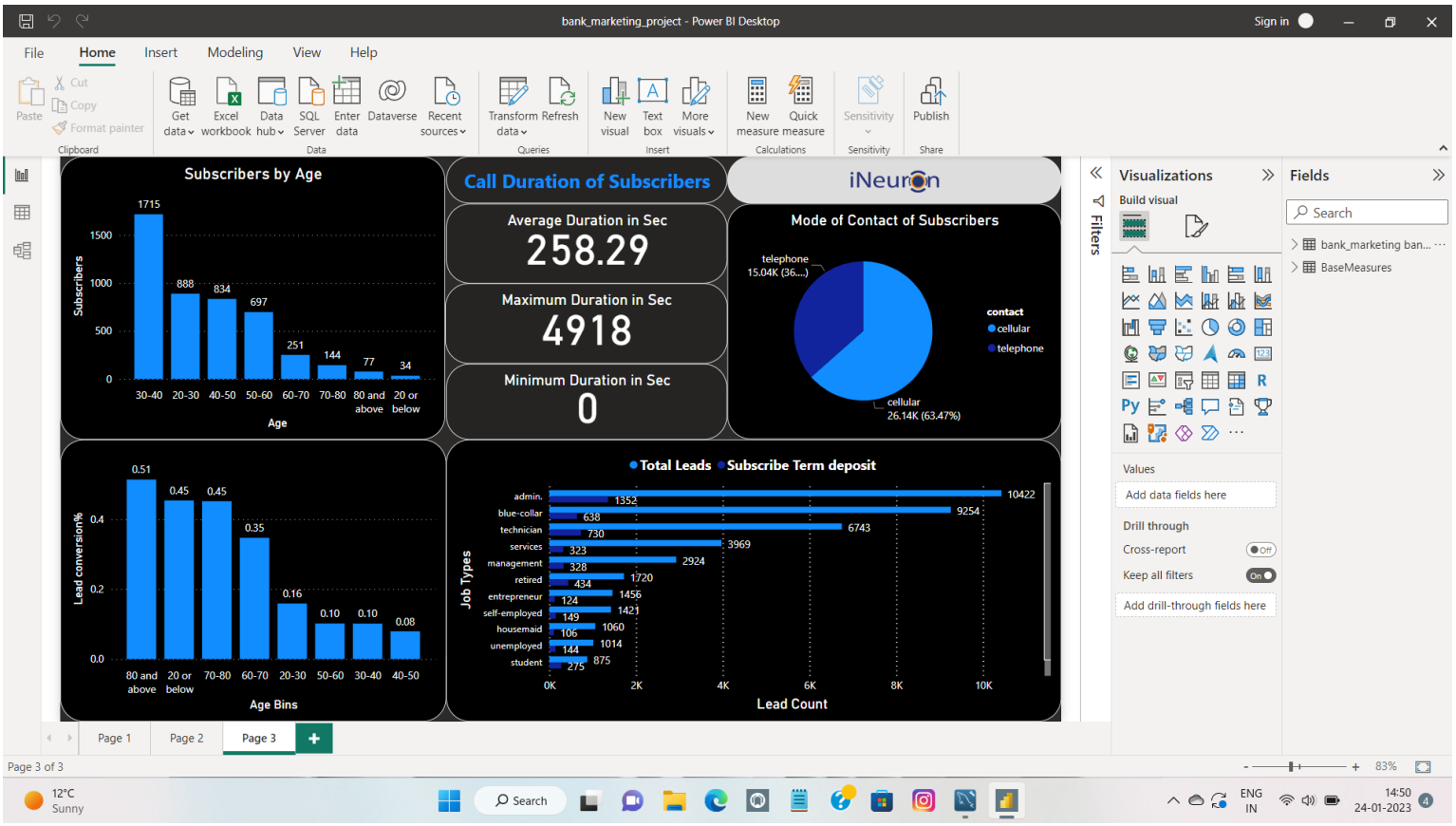
- There can be multiple rows and columns in the data.
- Each row represents a sample of data,
- Each column contains a different variable that describes the samples (rows).
- The data in every column can be a different type of data – e.g. numbers, strings, dates, Boolean etc.



3.1 Data to Insights through Visualizations and Excel Data Analysis


bank_marketing_project - Power BI Desktop
Bank Marketing Campaign Data Analytics
iNeuron
subscribers
no
yes
Subscribers Base
4640
Which type of jobs subscribe to term deposit most
Subscribers
Job Type
admin. 1352
technician 730
blue-collar 638
retired 434
management 328
services 323
student 275
self-employed 149
unemployed 144
entrepreneur 124
housemaid 106
unknown 37
Subscribers by marital
divorced 476 (10.26%)
single 1620 (34....)
married 2532 (54.57%)
marital
married
single
divorced
unknown
Subscribers by education
university.degree
high.school 1.03K
professional.course 0.60K
basic.9y 0.47K
basic.4y 0.43K
unknown 0.25K
basic.6y 0.19K
illiterate 0.00K
0.2%
Visualizations
Build visual
Fields
Search
bank_marketing bank_...
BaseMeasures
Column1
Conversion
Count of Ages of...
Count of jobs
Subscribers
Subscribers%
Total subscribers
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 2
Page 3
Page 1 of 3


bank_marketing_project - Power BI Desktop
Conversion by marital and y
marital
unknown
single
divorced
married
no 15.16%
yes 1...
no 14.1...
yes 14.15%
no 10.4...
yes 1...
no 10.2...
No of times contacted Subscribers Subscribers%
1 2300 0.50
2 1211 0.26
3 574 0.12
4 249 0.05
5 120 0.03
6 75 0.02
7 38 0.01
8 17 0.00
9 17 0.00
Total 4640 1.00
job Count of jobs Conversion
admin. 1352 0.13
blue-collar 638 0.07
entrepreneur 124 0.09
housemaid 106 0.10
management 328 0.11
retired 434 0.25
self-employed 149 0.10
services 323 0.08
student 275 0.31
Total 4640 0.11
iNeuron
Subscribers% by loan
loan
no
yes
unknown
yes 14.7...
no 82.97%
Conversion by month
Conversion
month
0.51
0.49
0.45
0.44
0.20
0.11
0.11
0.10
0.09
0.06
mar
dec
sep
oct
apr
aug
jun
nov
jul
may
age Count of Ages of subs Conversion
17 2 0.40
18 12 0.43
19 20 0.48
20 23 0.35
21 29 0.28
22 36 0.26
23 48 0.21
24 86 0.19
25 93 0.16
26 122 0.17
27 114 0.13
28 151 0.15
29 186 0.13
Total 4640 0.11
age Count of Ages Subscribers%
17 2 0.00
18 12 0.00
19 20 0.00
20 23 0.00
21 29 0.01
22 36 0.01
23 48 0.01
24 86 0.02
Total 4640 1.00
Visualizations
Build visual
Fields
Search
bank_marketing bank_...
BaseMeasures
Column1
Conversion
Count of Ages of...
Count of jobs
Subscribers
Subscribers%
Total subscribers
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Page 2
Page 3
Page 2 of 3



emp.var.rate	
Mean	0.081885501
Standard Error	0.007740691
Median	1.1
Mode	1.4
Standard Deviation	1.570959741
Sample Variance	2.467914506
Kurtosis	-1.062631525
Skewness	-0.724095549
Range	4.8
Minimum	-3.4
Maximum	1.4
Sum	3372.7
Count	41188

cons.price.idx	
Mean	93.57566437
Standard Error	0.002852156
Median	93.749
Mode	93.994
Standard Deviation	0.578840049
Sample Variance	0.335055802
Kurtosis	-0.829808577
Skewness	-0.230887652
Range	2.566
Minimum	92.201
Maximum	94.767
Sum	3854194.464
Count	41188

cons.conf.idx	
Mean	-40.50260027
Standard Error	0.022804816
Median	-41.8
Mode	-36.4
Standard Deviation	4.628197856
Sample Variance	21.4202154
Kurtosis	-0.358558311
Skewness	0.303179859
Range	23.9
Minimum	-50.8
Maximum	-26.9
Sum	-1668221.1
Count	41188

euribor3m	
Mean	3.621290813
Standard Error	0.008546254
Median	4.857
Mode	4.857
Standard Deviation	1.734447405
Sample Variance	3.0083078
Kurtosis	-1.406802622
Skewness	-0.709187956
Range	4.411
Minimum	0.634
Maximum	5.045
Sum	149153.726
Count	41188

nr.employed	
Mean	5167.01901
Standard Error	0.355647645
Median	5191
Mode	5228
Standard Deviation	72.178074
Sample Variance	5209.674366
Kurtosis	-0.015550986
Skewness	-1.041629312
Range	264
Minimum	4964
Maximum	5228
Sum	212819179
Count	41188