



# **ASSIGNMENT 2**

Qualification	BTEC Level 5 HND Diploma	BTEC Level 5 HND Diploma in Computing										
Unit number and title	Unit 14: Business Intelligence	Jnit 14: Business Intelligence										
Submission date		<b>Date Received 1st submission</b>										
Re-submission Date		Date Received 2nd submission										
Student Name	Nguyen Duc Cuong	Student ID	GCH18641									
Class	GCH0715	Assessor name	Doan Trung Tung									
Student declaration	l	1										

I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.

Student's signature

# **Grading grid**

Р3	P4	P5	P6	M3	M4	D3	D4





Summative Feedback:	<b>.</b>	Resubmission Feedback:
Grade:	Assessor Signature:	Date:
IV Signature:		





# TABLE OF CONTENTS

1		Intr	oduc	ction	6
2		Ove	ervie	w of Business Intelligence	6
	2.	1	Wha	at is Business Intelligence?	6
	2.	2	Busi	iness Intelligence process	6
3		Bus	iness	s Intelligence techniques	7
	3.	1	Coll	ection techniques	7
		3.1.	.1	Cleansing	7
		3.1.	.2	Labeling	8
	3.	2	Ana	lysis Techniques	9
		3.2.	.1	Dashboard	9
		3.2.	.2	Reporting	9
	3.	3	Ana	lytic Techniques1	0
		3.3.	.1	Regression1	0
		3.3.	.2	Machine learning1	0
4		Bus	iness	s Intelligence tools1	1
	4.	1	Prog	gramming tool (Jupyter)1	1
	4.	2	Prog	gramming Language (Python)1	2
	4.	3	Data	a Visualization tool (Tableau)1	3
5		Ехр	lain d	datasets1	4
	5.	1	Intro	oduction about Consultation to match couples1	4
	5.	2	Intro	oduction to datasets1	4
	5.	3	Som	ne problem in datasets1	5
6		Pre	-prod	cess on datasets1	5
7		Ana	alysis	Datasets_Design Dashboard1	7
	7.	1	Dasl	hboard 11	7
		7.1.	.1	The average age of drug use by men and women between religions1	7
		7.1.	.2	Incomes from work of drugs addicts1	8





	7.1.3	Locations of drugs addicts	19
	7.1.4	The age of having children of each orientation	20
	7.1.5	Evaluation Dashboard	20
-	7.2 D	ashboard 2	21
	7.2.1	Total income between men and women by the job	21
	7.2.2	The average age of each ethnicity	22
	7.2.3	Evaluation Dashboard	23
-	7.3 C	ashboard 3	24
	7.3.1	Distribution of religion by location and age	24
	7.3.2	Education affects income	25
	7.3.3	Evaluation Dashboard	26
8	Feedl	back results	27
9	How	Business Intelligence tools can contribute to effective decision-making	31
10	Legal	issues involved in the secure exploitation of business intelligence tools	31
	10.1	Too expensive and hard to justify the ROI of BI	31
2	10.2	Dealing with the impact of poor data quality	32
2	10.3	Poor functionality and interaction of Bl	32
2	10.4	Sluggish query and database performance	32
11	Conc	lusion	33
12	Refer	ences	34





# **List Of Figures**

figure 1: Jupyter	11
figure 2: Python	12
figure 3: Tableau	13
figure 4: Datasets	15
figure 5: The average age of drug use by men and women between religions	18
figure 6: Incomes from work of drug addicts	19
figure 7: Locations and drug addicts of men and women	19
figure 8: The age of having children of each orientations	20
figure 9: Dashboard 1	21
figure 10: Total income between men and women by the job	22
figure 11: The average age of each ethnicity	23
figure 12: Dashboard 2	24
figure 13: Distribution of religion by location and age	25
figure 14: Education affects income	26
figure 15: Dashboard 3	27
figure 16: Do you feel that using python for data cleaning based on our company's data is eaunderstand?	•
figure 17: Do you think that after clearing the data, the dataset is easier to track and better?	28
figure 18: Through this chart, do you think it has enough information for users to find the right ran you?	_
figure 19: Do you think this chart is informative and easy to understand to compare differences beincome by age and ethnicity?	
figure 20: Through this chart, do you think it has enough information for users?	30
figure 21: How much would you rate our presentation?	30





#### 1 INTRODUCTION

In this report a detailed overview of Business Intelligence is presented as well as the techniques in this area. Besides, it provides analyzing datasets along with the cleaning of data in datasets. From this data file, we applied BI to analyze and synthesize in dashboard. Finally, legal issues related to BI tools are presented in data analysis.

# 2 OVERVIEW OF BUSINESS INTELLIGENCE

### 2.1 WHAT IS BUSINESS INTELLIGENCE?

According to Pratt & Fruhlinger (2019) Business Intelligence (BI) leverages software and services to transform data into actionable insights that inform the organization's strategic and tactical business decisions. BI tools access and analyze datasets and present analytical findings in reports, summaries, dashboards, graphs, charts and maps to provide users with insights. about the state of the business.

For instance, Chris Hagans, vice president of operations at WCI Consulting, a consulting firm focused on BI, said: "A company that wants to better manage its supply chain needs the BI capability to identify it. where delays are happening and what do they exist during transit ". The company can also use its BI capabilities to discover which products are most often delayed or which modes of transport typically cause the most delays.

#### 2.2 BUSINESS INTELLIGENCE PROCESS

Business Intelligence is divided into **four different processes** (Cordell, 2019):

# Information gathering

- During the information collection phase, data is either prepared from existing sources (such as existing contact data, ERP data, financial databases, etc.) or is obtained from outside through Use live or online surveys, polls, questionnaires or forms.
- **Analysis:** Analysis is one of the key stages to turn data tho into information and make it easier for users to discover data:
  - Spreadsheet Analysis
  - Software that allows users to develop their own specific data queries
  - Visualization Tools

# 🖊 Reporting

 Reporting is the act of taking the analyzed data and presenting it in a way that creates a relationship between people or some sort of focus at which advantage will be gained through actions.





- Depending on the tools involved, reporting can occur as an extension of the analysis phase, but for the BI to be effective, the BI must be reported after being filtered or identified during the analysis phase. before being presented as a report.
- The report can be presented as a data sheet on screen or paper, but can also be displayed as a pivot table, chart or as executive summary in a company report.

# Monitoring and Prediction

- Monitoring allows users to track data and information in real time. The three main types of monitoring are:
  - **Dashboard**: Usually represented graphically to make it easier for the user to read.
  - **Key Performance Indicators (KPIs):** KPIs measure the performance of key drivers selected from the organization.
  - Business Performance Management: is a system designed to ensure that the performance goals for your organization or project are being met and that the results are being delivered.
- Prediction helps management predict what will happen based on existing data and other trends. There are two main types of predictions:
  - Data Mining: The act of finding patterns and relationships within and between large data sets to extract or convert data into something we can further understand and use.
  - Predictive Modeling: Any model designed to predict the outcome of an action or the probability of an outcome.

# 3 BUSINESS INTELLIGENCE TECHNIQUES

# 3.1 COLLECTION TECHNIQUES

# 3.1.1 CLEANSING

According to Eduardo Alonso (2019), Cleansing (or Data cleaning) is the process of detecting and/or removing corrupt or inaccurate records from a set of data. Data that negatively impacts your business is called dirty data. Cleansing includes the following steps:

- Identify corrupt data
- Correct the corrupt data
- Replace or delete inaccurate or incomplete parts of data.
- Advantages of data cleaning (Formplus Blog, 2020):





- Revenue Booster: Businesses who have the correct data on the demography of their target audience can employ the right marketing tactics. This will help generate more customers, sales, and higher revenue.
- Cost-effective: When working with the right database for marketing, businesses are sure of getting a high engagement rate, giving back the required value for their money. This will help save costs spent on ineffective marketing practices.
- o *Increases Productivity*: With accurate and updated information, employees will spend less time contacting expired contacts or customers with stale information. For example, if support tickets are not updated when completed, employees will waste time contacting customers when they do not need to.
- Boosts Reputation: Having clean and error-free data will help boost trust and reputation, especially for companies that specialize in sharing data with the public. If you provide clean data to people, they will trust you as a reliable data bank.

# ➡ Disadvantage of data cleaning (Formplus Blog, 2020):

- Analysts may lose out on actionable insights due to incomplete data. This is very common in cases where missing observations and outliers are dropped.
- o It may lead to an even bigger problem when automated. Some automated data cleaning tools are not very smart and may end up mishandling some observations in the dataset.
- It is time-consuming. Data cleaning may take a lot of time, especially when dealing with large data.
- o The process is very expensive.

#### 3.1.2 LABELING

According to Margaret Rouse (2019), Labeling (or Data labeling) is the process of detecting and tagging data samples. The process can be manual but is usually performed or assisted by software. Data labeling offers labeling for input and output data classification, creating a learning foundation for future data processing.

# **Advantages of data labeling** (Margaret Rouse, 2019):

- You can correctly identify the information in the order in which the labels are shown after studying the input and output data.
- o You can discover similarities between details to see differences from other data.
- It is an iterative process, so as more knowledge is gathered, it is scalable and agile.

# **♣ Disadvantages of data labeling** (Margaret Rouse, 2019):

- o It is time consuming and expensive to implement manual data labeling.
- Labeling makes work riskier.





# 3.2 ANALYSIS TECHNIQUES

#### 3.2.1 DASHBOARD

According to tableau.com (2020), Business intelligence dashboards (BI dashboard) are solutions for data management and data visualization used for analyzing your data. Interactive features such as filters and actions can be used by content developers to merge charts, graphs and reports on a single screen for snapshot overviews.

Dashboards are one of BI platforms' most common features because they provide easily understandable data analysis, allow you to customize what data you want to view, and provide a way to share your analysis results with others. BI dashboards are now an important component of the leading platform for cloud-based analytics.

# **Advantages** (Melanie Chan, 2019):

- o **Enhanced Visibility**: BI Dashboard provides better visibility with available information whenever required to ensure businesses are better positioned to respond to visual conditions field changes.
- Save time efficiently: With the BI dashboard you no longer waste valuable time generating reports from multiple systems. Instead, the data is taken from a centralized source and displayed as an easy-to-understand visual overview.
- o **Better forecasting**: Businesses can more effectively plan for fluctuations in demand for the next business cycle, set measurable and deliverable goals for greater success.

# ♣ Disadvantages (Melanie Chan, 2019):

- Flashy or cluttered design, in which users try to combine too much information without understanding constraints or consider their specific needs from different scopes of detailed, measurable data analysis provided.
- The difficulty in attaching the support data to the dashboard, and the data cannot be refreshed automatically means that both of these tasks have to be done manually.

# 3.2.2 REPORTING

Reports are a fundamental part of business intelligence, focusing on visualizing data in different types of visualizations like tables, graphs, and charts. The visualization in the reporting context is a graphical representation of the data, whose goal is to accurately present information in a digestible form to the end user.

Advantages of reporting technique (Sandra Durcevic, 2020):





- Speed up workflow: Each information can be gathered into a single, live dashboard, which will ultimately ensure a fast, clear, simple, and efficient workflow. This type of report will become intuitive, accessible and stable in gathering insights.
- Performance optimization and forecasting: Business intelligence and reporting not only focus on tracking, but also include forecasting based on predictive analytics and artificial intelligence that can easily help avoid decision-making. Expensive and time-consuming business planning.
- Cost optimization: Using business reporting software, costs and savings of most businesses will be measured in an optimal and fastest way.

# 3.3 ANALYTIC TECHNIQUES

# 3.3.1 REGRESSION

According to Sunil Ray (2016), Regression analysis is a form of predictive modelling technique which investigates the relationship between a dependent (target) and independent variable (s) (predictor). This technique is used for forecasting, time series modelling and finding the causal effect relationship between the variables. There are 7 types of regressions:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Stepwise Regression
- Ridge Regression
- Lasso Regression
- ElasticNet Regression

# Advantages

- For a small business, it helps to determine which factors are most important, what can be ignored and how they interact.
- Provides a powerful statistical method that allows firms to examine relationships between two or more variables of interest.

# Disadvantages

- It involves a very long and complex computational and analytical process.
- o It cannot be used in the case of the qualitative phenomena of honesty, crime, etc.

#### 3.3.2 MACHINE LEARNING





According to Expert System Team (2020), Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

The purpose of machine learning is to allow computers to automatically learn and adjust actions accordingly without human intervention or assistance.

# Advantages

- There is no need for human interference because machines are learning, enabling them to make predictions and improve their algorithms themselves as well.
- Large volumes of data can be reviewed and specific trends and patterns that are not clearly visible to humans can be uncovered.

# Disadvantages

- It takes a large of resources to operate.
- High probability of errors.

# 4 BUSINESS INTELLIGENCE TOOLS

# 4.1 PROGRAMMING TOOL (JUPYTER)

According to Computerhope.com (2020), Jupyter, formally known as the Jupyter Project, is a non-profit organization that creates open-source interactive computing software. It was founded in 2014 by Colombian physicist and software developer Fernando Pérez, as a language-agnostic version of the IPython (interactive Python) project.



figure 1: Jupyter





Jupyter includes tools to interactively write code, display real time results, and publish results on the web for mathematicians, scientists, teachers, data analysts, and computing professionals.

# Advantages (Rajesh Sampathkumar, 2019):

- Autocomplete and quick function help lookup
- Doing visualizations within the IDE has never been easier
- Can be run as a service remotely and opened on a client browser
- Documentation can be done right there along with the code
- Presentation mode
- Easy collaboration

# ♣ Disadvantages (Rajesh Sampathkumar, 2019):

- o In-memory variables can be overwritten
- Tool support for notebook files (such as ipynb) can vary Gitlab provides decent support but Bitbucket, for instance, does not even parse the JSON.

# 4.2 PROGRAMMING LANGUAGE (PYTHON)

According to techterm.com (2010), Python is a high-level programming language designed to be easy to read and simple to implement. It is open source, which means it is free to use, even for commercial applications. Python can run on Mac, Windows, and Unix systems and has also been ported to Java and .NET virtual machines.



figure 2: Python

# Advantages (techVidan.com, 2020):

- Easy to read, learn and write because it needs less lines of code to perform the same task as compared to other major languages like C/C++ and Java.
- Improved Productivity because Python is a very productive language. Due to the simplicity of Python, developers can focus on solving the problem.
- The standard library of Python is huge, you can find almost all the functions needed for your task.





- Python is an interpreted language which means that Python directly executes the code line by line. In case of any error, it stops further execution and reports back the error which has occurred.
- Python shows only one error even if the program has multiple errors. This makes debugging easier.

# Disadvantages (techVidan.com, 2020):

- o Slow speed because it has to do the extra work while executing code.
- The Python programming language uses a large amount of memory. This can be a disadvantage while building applications when we prefer memory optimization.
- This can be a disadvantage while building applications when we prefer memory optimization.

# 4.3 DATA VISUALIZATION TOOL (TABLEAU)

Currently, there are many data visualization tools as Tableau, Power BI, Google Analytics, Domo, etc. The tool our team has chosen for visualization of the data is the Tableau software.

According to (tableau.com, 2020), Tableau is a visual analytics platform transforming the way we use data to solve problems—empowering people and organizations to make the most of their data.



figure 3: Tableau

- **♣** Tableau also offers three standalone products:
  - Tableau Desktop (for anyone)
  - o Tableau Server (analytics for organizations), which can be run locally
  - Tableau Online (stored analytics for organization)
- Advantages (KnowledgeHut, 2019):
  - High Performance: Users rate Tableau's overall performance as strong and secure. It can handle millions of rows of data with ease. The huge advantage of having Tableau is different types of visualization can be created at one shot.





- Easy to upgrade: Tableau customers are happy with using the latest release of the software because the upgrades are easy to be carried out.
- Ease of use: It is easy to use since it is simple user interface software. Also, it is simple to drag and drop interface which is very easy to learn.
- Mobile-Friendly: There is an accomplished mobile app available for IOS and Android which adds mobility to Tableau users and allows them to keep statistics at their fingertips. The app supports practically that a desktop and online version has.

# Disadvantages (KnowledgeHut, 2019):

- Poor Versioning: The main disadvantage of using Tableau is, only recent versions supports revision history and for the older one's package rolling back is not possible.
- Need manual effort: Tableau's parameters are inactive and only a single value can be selected using a parameter. You need to update it manually whenever the data gets change.
- No version control: Once the dashboards and reports are published on the server you cannot get back to the previous levels of data in Tableau. It is not possible to go back and recover old data.

# 5 EXPLAIN DATASETS

#### 5.1 INTRODUCTION ABOUT CONSULTATION TO MATCH COUPLES

**OkCupid** is a mobile dating app that aims to register and match users to arouse a potential romance. The purpose of this app is to create an environment where single people can find partners based on cooperation. This is the basis for building an effective relationship involving two people:

- When you people whose hearts are set on similar things:
- Their minds work together
- Their plans are the same
- Their objectives are the same.

#### 5.2 INTRODUCTION TO DATASETS

The OkCupid profile data set contains information from 59,946 San Francisco users. This data set has been collected from activity records containing at least one photo as of June 2012 and is made available to the public. Each profile contains information about sex, sexual orientation, ethnicity, religion, education, physical characteristics, and more out of 31 possible traits. Income, height, and age are the only usable numeric characteristics, making most of the data categorical. Ten possible essays are also available, answering prompts like "Describe your typical Friday night" and "What's the first thing people notice about you?"







Age	status	sex	prientation	body_type	rilet	drinks	rings	education	ethskity	height	income	joh	last online	location	offspring	pets	religion	sign
22	single	m	straight	a little extra-	strictly anything	socially	never	working on college	Fasier, white:	75	4	transportation	2012-06-28-20-30	south san francisco,	doesn't have kids, b	Nes dogs and likes	agreeticism and	ven gemini
35	zingle	m	straight	аналада	mostly other	aften	sometimes	working on space of	aiwhite	70:	80000	hospitality/travel	2012-06-29-21-41	pakland, california	doesn't have kids, b	likes degs and likes	agrodicte but r	of cancer
4 38	avelable	m	straight	thin	anything	socially		graduated from ma	6	68	-1.		2012-06-17-05-10	san francisco, califlo	6	has oits		pinces bu
5 23	single	m	straight	this	regetarian	socially		working on college	/white	72	20000	student.	2012-06-28-14-22	betkeley, california	doesn't went kids	Titles cats		prom
6. 77	single	m	straight	athletic		antially	reset	graduated from col	Design, black, other	65	-1	artistic/musical/w	2012-00-27-21-28	san francisco, califo		likes dogs and likes	1.	aquartus
7 29	single	m	straight	ave/age	mostly anything	ascially		graduated from col	lewhite:	67	-1	computer / hordway	2012-06-29-19-18	san francisco, calific	doesn't have kids, b	rikes arts	atheorn	taurus
0 32	single	4	straight	#t	strictly anything	totally	rever	graduated from col	livwhite, other	65	-1		2812-96-25-29-45	san francisco, califo	t	likes dogs and likes	4	virgo
9 31	single	4	straight	анагодя	mostly anything	seculty	never	graduated from col	licwhite	65	-1	artistic/musical/w	2012-06-29-12-30	san francisco, califo	doesn't have kids, b	likes dogs and likes	emodunity	sagmanu
10 34	single	1	straight		strictly anything	socially		graduated from col	lkwhite .	67	4		2012-06-29-29-39	belvedere tiburon,	c doesn't have kids	likes dags and likes	ichristianity but n	ot t geraini bu
11:12	single	m:	straight	athletic	mostly anything	notatali	rever	working on two-ye	arwhite:	65	-1	student	2012-06-28-31-08	san mateo, californ		likes dogs and likes	ratheiom and loug	hin cancer bu
2 25	available	m	streight	average	reastly anything	sensity			white	70	-1		2012-00-04-14-59	san franceco, califo	(			taurus
7 28	seeing someone	m	straight	awerage	mostly anything	socially	rever	graduated from col	lkwhite.	72	40000	banking / financial /	2012-05-22-17-05	daly city, california		likes orb	christianity and v	ery leo but it
4 34	single	m	straight			often			white	72	1	entertainment/me	2012-05-28-21-28	san francisco, califo	r doesn't have kids		other	teores
5 30	single	1	straight	Stoney	mostly anything	unially	never	graduated from hig	tiwhite	66	30000	sales / marketing / t	2012-06-13-16-06	san francisco, califo		has dogs and likes o	s christianity but o	oft
6 29	single	1	straight	the	mostly anything	socially	never	working on college	/inspanic/latin, whi	162	30000	other	2012-06-29-08-55	san leandro, califon	ndoesrf have kids, b	Wes dogs and has o	acathelisses	tuns
17 39	single	1	straight	fit	strictly anything	socially		graduated from col	lkwhite:	65	-1	nedicine/health	2012-06-29-22-09	san francisco, califo	r doesn't have kids	likes dogs and has o	siathelon and laug	hin aquartus 3
18 11	single	m	draight	fit		socially		working on master	5 While	30	4	entertainment / me	2512-06-29-16-00	switand, railfornia		likes dogs and likes		pisces but
19:26	single	#	straight	average		rectally	rever	working on college	/ihispanic/latin	64	-1	banking/financial/	2012-06-29-21-17	sen francisco, calific	doesn't have kids			artes and
20 71	single	m	straight	average	mostly vegetarian	rarely	reser	graduated from spi	cwhite	п	-1	medicine / health	2012-16-29-04-04	san francisco, califo		likes dogs and likes	ichnistianity	None but r
El 33	single.	m	straight	athletic	mostly anything	socially	reset	graduated from me	s white	72	-1	tolence/tech/eng	2512-06-27-21-41	san francisco, califo	doesn't have kids	likes dogs and likes	catholicism but n	ot falloes and
22 27	single	1	straight	average	mostly anything	socially	heset	graduated from col	licwhite	67	-1	sales / marketing / t	2012-06-26-20-40	san francisco, califo	£	likes dogs and dislik	ocatholicism	Hbra
22	single	1	straght	athletic		socially	reset	working on college	Ø.	67	-1	student	2012-06-15-22-56	san francisco, califo	rdoesn't have kids, b	likes degs and likes	rathelissn and s	on tauns bu
14,30	single	m	straight	fit	mostly anything	senally	rever	graduated from col	licwhite:	69.	-1	esecutive / munage	(2012-06-38-15-22	san francisco, califo	r doesn't have kids, to	likes dogs and likes	ragnosticism and r	en sagithiriu
25 30	single	m	streight	thre	strictly vegon	notatali	never	graduated from col	Exwhite	n -	-1	education / academ	(2012-05-17-28-2).	sen francisco, califlo		likes dogs and likes	ratherm and long	hinaquirius
33	single	m:	straight	then	strictly anything	secially	sometimes	graduated from col	lkwhite	33	-1	ertistic/musical/w	2012-06-29-21-32	pakiend, california		Westorb	other .	зсограз аг
27, 28	single	m	straight	ht.	anything	rarely	rever	graduated from sol	Rossian, white:	Ti	-1	neditire / health	2012-06-26-01-27	atheton, ralifornia	ploesn't have kids, b	nhas dogs		gerain an
29 22	single	m	straight	fit	mostly anything	socially	never	working on college	Figacific islander, wh	172	-1	clerical / administra	2012-06-30-01-17	san francisco, califo		likes dogs and doll	catholioses and b	augleo and it
29 22	zingle	m	straight	average	anything	socially	19990		2537	67	4	other	2012-06-25-23-06	oakland, california		has dogs and dislike	agnoction and i	laugileo and it
10 10	single	m	straight	fit		socially		graduated from ma	is .	34	-1		2012-06-19-05-22	san francisco, cáliflo	6	lites dogs and lites	rathelon and laug	his
11 12	single	m	straight	70		seculty	sometimes	graduated from col	Kwhite	65	-1	other	2012-06-29-07-01	san francisco, califo	6		agnosticum and i	Handowijani
27	single	1	straight	average.	anything	notially		working on college	/iwhite	64	-1	other	2011-11-10-11-15	san francisco, califo			agnorticism	genin
B 27	single	m	straight	athletic	anything	seculty	reser	graduated from col	le .	72	-1	science/tech/eng	2012-06-27-29-16	san francisco, calific	doesn't have kids, b	ilkes dogs	atheorn	
34 38	single	*	straight	анегоде	mostly anything	possibly	rever	graduated from col	l/white	67	-1	computer / hardwar	2012-05-25-00-04	san leandro, califor	doesn't have kids		christianity	Hora
35 20	single	*	straight	average	mostly vegetarian	seculty	never	working on college	Vi .	60.	-1	hospitality / travel	2012-06-30-00-00	can rafael, california	doesn't want kids	has dogs and likes o	1	cancer
10.11	- similar	100	mode	484.0	market in display	models.	deint.	mand arteral finding and	South train	ça.	1	- and and assessment Con-	2012/06/2014	medianological soldier		Street Agent and Street	official radios	Annual of Street, or

figure 4: Datasets

# 5.3 SOME PROBLEM IN DATASETS

OkCupid data file is a lot and then there was a lot of grain, making the data not optimal. Below are a few issues that I give a remedy to clean up data.

- In Sex column: Change the data "m, f" to "Male, Female"
- ♣ In Height column: Change data to Integer.
- In Income column: Change negative number to 0
- In Job column: Change NULL data to "Unemployed"
- In Offspring column:
  - Change "Has Kids" data to "Yes"
  - The remaining data changes to "No"
- In Pets column:
  - Convert "Like and have" animal data to "Yes"
  - The remaining data changes to "No"
- In Religion column: Change NULL data to "No"

# 6 PRE-PROCESS ON DATASETS

During development we use Python to clean up data.





♣ Step 1: We add the required libraries: The CSV library is used to read and write CSV files.

# import csv import numpy as np

**Step 2**: *In Sex column*: Change the data "m, f" to "Male, Female"

```
# Change m, f to male, female
def changeSexName(row):
   'Male' if row[2] == 'm' else 'Female'
   return row
```

♣ Step 3: In Height column: Change data to Integer

```
for row in datas:
    if(row[10] == ''):
        row[10] = 0

else:
    row[10] = row[10].split('.')[0]
```

```
# Change height to integer
def fixNumbers(row):
    row[10] = int(row[10])
return row
```

**♣ Step 4**: *In Income column*: Change negative number to 0

```
# Income: Change negative number to 0
def changeNumber(row):
   if row[11] == "-1" or row[11] == "":
      row[11] = 0
   else:
      row[11] = int(row[11])
   return row
```

♣ Step 5: In Job column: Change NULL data to "Unemployed"

```
# Job: Change NULL to "Unemployed"
def changeJobName(row):
   if row[12] == "":
       row[12] = "Unemployed"
   return row
```

- Step 6: In Offspring column:
  - The data that means having children will change to "Yes"





The remaining data changes to "No"

```
# OFFSPRING: has kid => yes, doesn't => no
def offSpring(row):
   if row[15] == "has a kid" or row[15] == "has kids":
      row[15] = "Yes"
   else:
      row[15] = "No"
   return row
```

- Step 7: In Pets column:
  - Convert "Like and have" animal data to "Yes"
  - The remaining data changes to "No"

```
# Pets: liked => YES else NO
def changePets(row):
    'No' if row[16] == '' else 'Yes'
    return row
```

♣ Step 8: In Religion column: Change NULL data to "No"

```
# Religion: Change Null => NO
def changeReligion(row):
   if row[17] == "":
      row[17] = "No"
   return row
```

♣ Step 9: In these columns, we will delete the blank data to make the entire data look cleaner and better.

♣ Step 9: In these columns.

■ The step 9: In the st

```
# remove null data
datas = list(filter(lambda row: True if row[4] != '' else False, datas))
datas = list(filter(lambda row: True if row[5] != '' else False, datas))
datas = list(filter(lambda row: True if row[6] != '' else False, datas))
datas = list(filter(lambda row: True if row[7] != '' else False, datas))
datas = list(filter(lambda row: True if row[8] != '' else False, datas))
datas = list(filter(lambda row: True if row[9] != '' else False, datas))
datas = list(filter(lambda row: True if row[12] != '' else False, datas))
datas = list(filter(lambda row: True if row[15] != '' else False, datas))
datas = list(filter(lambda row: True if row[16] != '' else False, datas))
datas = list(filter(lambda row: True if row[17] != '' else False, datas))
datas = list(filter(lambda row: True if row[18] != '' else False, datas))
datas = list(filter(lambda row: True if row[19] != '' else False, datas))
```

# 7 ANALYSIS DATASETS DESIGN DASHBOARD

# 7.1 DASHBOARD 1

#### 7.1.1 THE AVERAGE AGE OF DRUG USE BY MEN AND WOMEN BETWEEN RELIGIONS

This chart shows the average age of drug addiction among both men and women in all religions. In the chart:





- Blue represents the average age who never uses drugs
- Orange represents the average age who occasionally uses drugs
- Red represents the average age who regularly uses drugs.

Through this chart we can determine what is the average age of drug use in all religions. From there, it is possible to determine the percentage of drug use and non-drug use at all ages of religions.

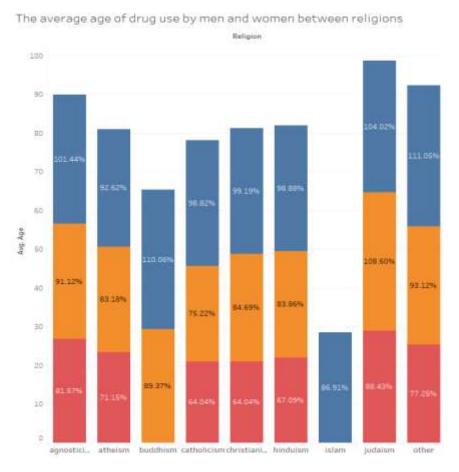


figure 5: The average age of drug use by men and women between religions

# 7.1.2 INCOMES FROM WORK OF DRUGS ADDICTS

This chart shows how much people earn from using drugs and not using them by job type. From the chart we can determine that people who do not use drugs earn much more than those who use them at any job.





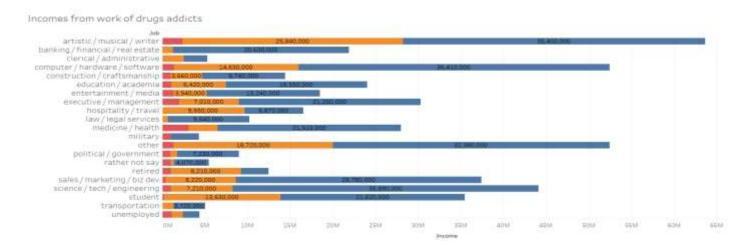


figure 6: Incomes from work of drug addicts

#### 7.1.3 LOCATIONS OF DRUGS ADDICTS

This chart shows the exact mean age of women and men based on their location, and is similar to the charts above showing the average age of drug users and non-users separately. Based on this chart we can see what each region will have the average age of men and women.



figure 7: Locations and drug addicts of men and women





#### 7.1.4 THE AGE OF HAVING CHILDREN OF EACH ORIENTATION

This is a very simple chart. This chart shows the percentage of people having children based on their orientation. We can determine what the average age by each orientation is the percentage of children.

- Red represents "Have no kid"
- Green represents "Have Kids"

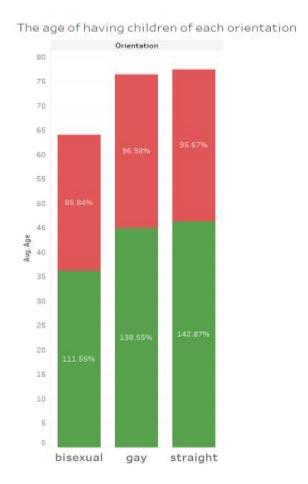


figure 8: The age of having children of each orientations

#### 7.1.5 EVALUATION DASHBOARD

The dashboard aggregates the data of everyone using Okcupid and focuses on those who use drugs according to their religion. Besides, it also shows where their location is. At the same time, it also shows how much their income is by each job and evaluates the average percentage of age with children of people according to each orientation.

From this information, OkCupid users can get general information on each region of drug users and not using it and their income. From there, OkCupid can filter out the objects that suit your needs.









figure 9: Dashboard 1

# 7.2 DASHBOARD 2

# 7.2.1 TOTAL INCOME BETWEEN MEN AND WOMEN BY THE JOB

This chart shows the total income between male and female by job type. Total income of each type of job for male, the total income of each type of job for female. From that, we can see which job has the highest gross income and which job has the highest gross income for males and females of all jobs. Blue represents the female and orange represents the male. When choosing a job, you can look at this chart to choose the job with the highest gross income, so that you can decide which job is right for you.





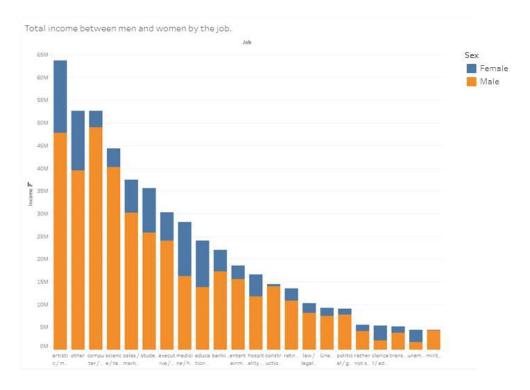


figure 10: Total income between men and women by the job

# 7.2.2 THE AVERAGE AGE OF EACH ETHNICITY

This chart shows the average age of both males and females of each ethnicity. This allows you to compare and analyze which ethnicity has the lowest and highest mean age and from that result, we can know which ethnic group's young people usually gather.





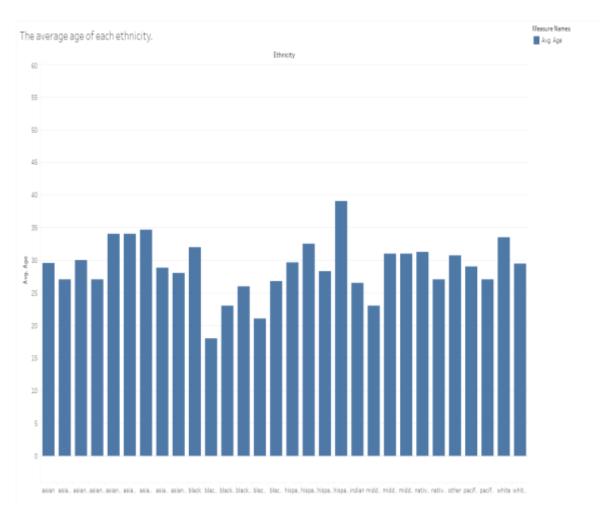


figure 11: The average age of each ethnicity

# 7.2.3 EVALUATION DASHBOARD

This chart is a combination of two related charts. This chart helps you see the average age of male or female in each ethnic group with high gross income for each job they do.

In conclusion, we can compare the difference in gross income by ethnicity and age. From there, it is possible to evaluate the jobs that bring in the highest income for each ethnic group.





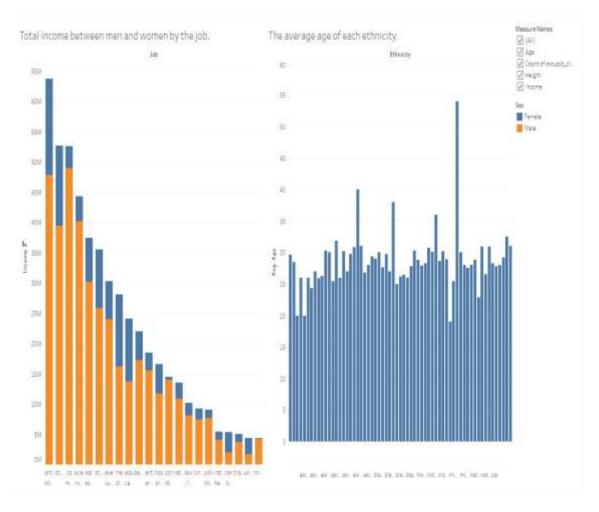


figure 12: Dashboard 2

# 7.3 DASHBOARD 3

# 7.3.1 DISTRIBUTION OF RELIGION BY LOCATION AND AGE

The chart shows the distribution of locations through religion and the population concentrated in that area. It can be seen that the San Francisco area is the most densely populated, the number of religions there is also higher than in other areas, especially the non-religious people are very concentrated in this area.





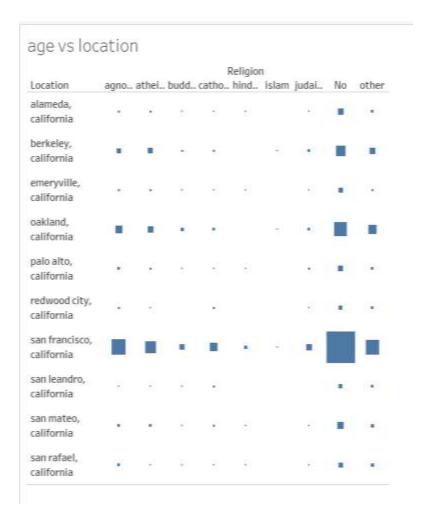


figure 13: Distribution of religion by location and age

# 7.3.2 EDUCATION AFFECTS INCOME

Based on the graph, it can be seen that people graduates from college or university have very high incomes compared to those who drop out of college in the middle or just graduate from high school. Total income among men is 2 or 3 times higher than woman.





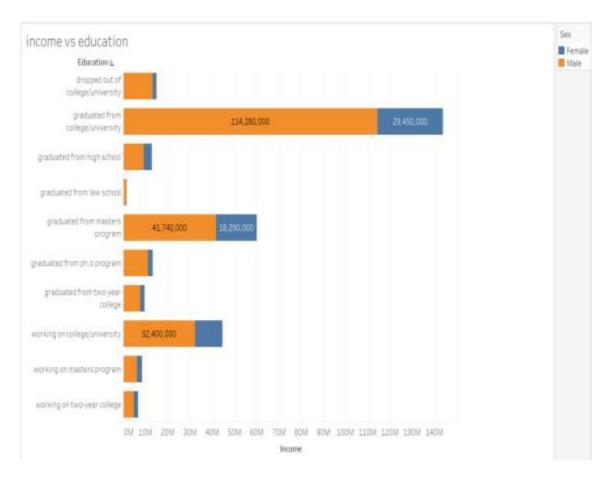


figure 14: Education affects income

# 7.3.3 EVALUATION DASHBOARD

Based on the graph, it can be seen that people graduates from college or university have very high incomes compared to those who drop out of college in the middle or just graduate from high school. Total income among men is 2 or 3 times higher than woman.







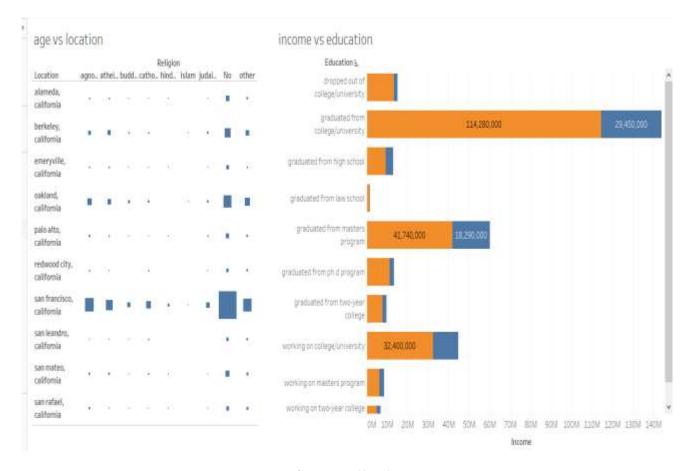


figure 15: Dashboard 3

# 8 FEEDBACK RESULTS

My team collected data using Google Forms To evaluate the results from group presentations with GCH0715 class members and faculty. The questions are built to comment on and evaluate my team's presentation such as BI overview, using Python for data set deletion and dashboard design and data analysis.

**♣** Do you feel that using python for data cleaning based on our company's data is easy to understand?

After using Python to clean up data. Up to 80% of users responded that it is simple to understand, besides only 20% of users are wondering but we can see no one responded that it is confusing.





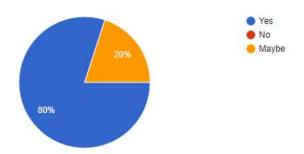


figure 16: Do you feel that using python for data cleaning based on our company's data is easy to understand?

♣ This is our dataset. OkCupid is a mobile dating app that aims to register and match users to arouse a potential romance. Do you think that after clearing the data, the dataset is easier to track and better?

Through the survey results, we can see that more than half of the users, specifically 62.5% agree that after cleaning the data file, this profile is easier to analyze than before.

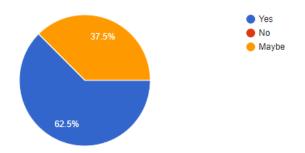


figure 17: Do you think that after clearing the data, the dataset is easier to track and better?

♣ The dashboard aggregates the data of everyone using Okcupid and focuses on those who use drugs according to their religion. Besides, it also shows where their location is. At the same time, it also shows how much their income is by each job and evaluates the average percentage of age with children of people according to each orientation. Through this chart, do you think it has enough information for users to find the right range for you?

We can see, in the first dashboard, all user feedbacks agree that this dashboard has provided enough information along with analysis data to help users find the right audience for themselves. This dashboard has shown the relationship between the use of drugs affecting people's income and the work of people in religions in each region.







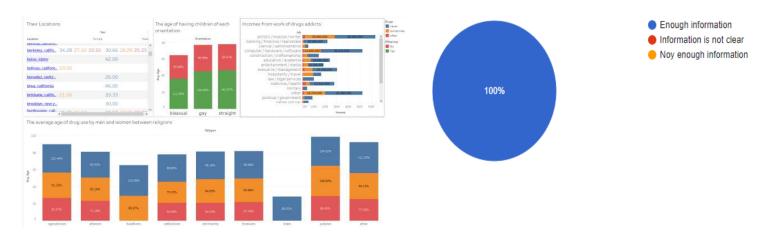


figure 18: Through this chart, do you think it has enough information for users to find the right range for you?

This chart is a combination of two related charts. This chart helps you see the average age of male or female in each ethnic group with high gross income for each job they do. In conclusion, we can compare the difference in gross income by ethnicity and age. From there, it is possible to evaluate the jobs that bring in the highest income for each ethnic group. Do you think this chart is informative and easy to understand to compare differences between income by age and ethnicity?

In the second dashboard, 87.5% of users agreed with the survey results that this dashboard has analyzed the complete data. However, up to 12.5% of users respond that this information is not enough to raise the issue we are talking about. In short, from the results, we still know that this dashboard isn't too bad.

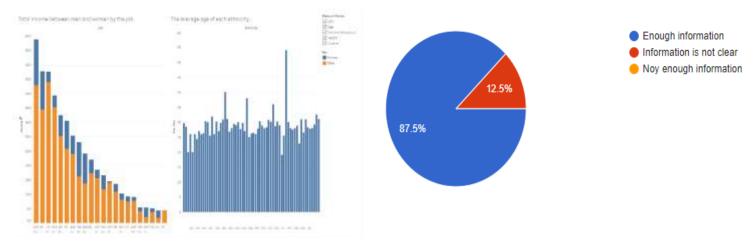


figure 19: Do you think this chart is informative and easy to understand to compare differences between income by age and ethnicity?

Based on the graph, it can be seen that people graduates from college or university have very high incomes compared to those who drop out of college in the middle or just graduate from





high school. Total income among men is 2 or 3 times higher than woman. Through this chart, do you think it has enough information for users?

In this last dashboard, all user responses told our team that the dashboard has analyzed all the data to show total income across regions by age. This can help us see which areas have high levels of income and high population literacy.

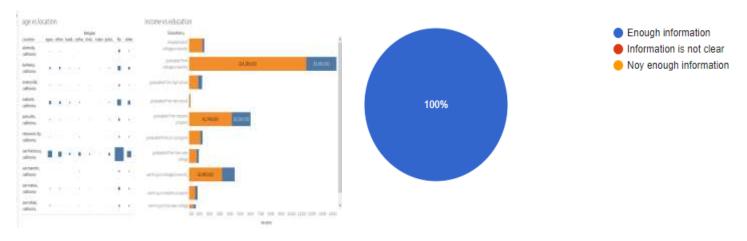


figure 20: Through this chart, do you think it has enough information for users?

# How much would you rate our presentation?

From the question of objective assessment of the group's presentation to draw the strengths and weaknesses of the group members in terms of content and presentation. Data collected from the class members, The results include: 50% rating 5 points, 37.5% rating 4 points, and finally 12.5% rating 3 points on a scale. 5. From this collected information, the whole group also completed the article with good results, but our team learned from experience to improve the content as well as improve the presentation and teamwork skills as well. Presentation skills in English, analytical skills, and explanations to be further improved after the project

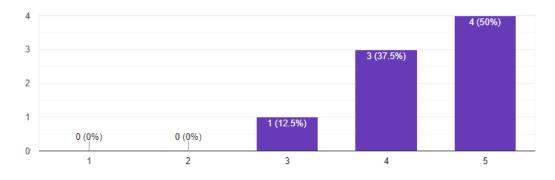


figure 21: How much would you rate our presentation?





# 9 HOW BUSINESS INTELLIGENCE TOOLS CAN CONTRIBUTE TO EFFECTIVE DECISION-MAKING

**Example 1**: The decision making to re-locate head office

BI tools help companies do financial analysis in surrounding areas. Evaluate whether the company's budget is appropriate for the rent in the new area. Moreover, the BI tool helps the company to control money better. It is also helped the company analyze the geographic area, facilities, and living standards of the people there. Based on that, find appropriate locations for new facility transfers can be drawn.

BI tools help companies analyze the market in that area. Use analytical tools such as Tableau to analyze clients' metrics, sexual orientation, their total income in that area, and the distribution of the population in those regions. All collected data can be easily translated into actionable information, which will make it easier for the company to select new locations.

**Example 2**: Decision making to attract business advertising strategy

BI tools help the company analyze the right customers for the company based on criteria such as age, marital status, habits, ... so that the company can make decisions about suitable advertising strategy.

Use Bi to analyze advertising strategies of other companies, the social media platforms they use, and automated posters in the mall. Based on that, the suitable strategy can be drawn to compete with other companies on social networks. Leads to better business decisions and outstanding financial performance. Help the company have a more perspective view of the strategy than follow their own intuition.

# 10 LEGAL ISSUES INVOLVED IN THE SECURE EXPLOITATION OF BUSINESS INTELLIGENCE TOOLS

#### 10.1 TOO EXPENSIVE AND HARD TO JUSTIFY THE ROLOF BI

The pricing of the BI implementation is of the utmost concern for small and medium-sized enterprises (SMEs). In addition, limited resources make it impractical and worrisome to find qualified professionals such as data science specialists, IT infrastructure specialists, and consultant analysts. In the past, expensive enterprise BI solutions required huge hardware resources. In addition to the increase in deployment costs, the establishment of these data warehouses and processors is also impacting expensive IT labor (Durcevic, 2019).

**♣ Solution**: A strong argument for BI ROI, agile, innovative and cloud BI software solutions such as datapine help small businesses to collect the data they need without needing anyone to spend significant quantities of money to implement. Small businesses can also select BI suppliers that provide flexible solutions that do not need heavy IT support and can be provided at the right price.





# 10.2 DEALING WITH THE IMPACT OF POOR DATA QUALITY

In today's digital world, there is so much data available, so one of the most challenging business issues today is the filtering and exploitation of gold data that will help you boost your business performance. With modern data volumes linked to BI, there are two obvious problems (Durcevic, 2019):

- Critical information is deeply hidden and is often lost or skipped in systems, software and platforms.
- Bad, unreliable, and complex data-driven observations are provided by BI-centric systems and services, eventually wasting time and money while making incredible progress.
- **♣ Solution**: You need strategies to drive quality management through the organization to make sure that you avoid negative data. In addition to working with dynamic KPIs that match your business's unique needs, goals, and initiatives, you should also implement a Data Quality Management (DQM) approach and support those that follow suit in business.

# 10.3 POOR FUNCTIONALITY AND INTERACTION OF BI

This issue is faced by many businesses because people need to gather, process and interpret data from a variety of sources, mostly from highly hand-crafted platforms or programs and without interactivity. Weak processes and functions more often than not lead to inaccuracies, inefficiencies, and expensive errors that can potentially hold back a business (Durcevic, 2019).

**♣ Solution:** By working with interactive BI dashboards, you'll be able to go beyond static reporting tools like Excel, PowerPoint, and other static reporting tools and have access to visual data instead. Interactive, intuitive will help you achieve worthy results.

# 10.4 SLUGGISH QUERY AND DATABASE PERFORMANCE

Slow database performance, one of the most commonly publicized BI challenges to date, leads to a range of internal challenges, including security problems with business intelligence, fragmented processes, weak interdepartmental communication, and severe reporting lags. None of these problems are good for company (Durcevic, 2019).

# Solution:

- It is important to not only monitor the quality of your data and prepare it thoroughly for analysis in order to drive productivity and overcome this common BI challenge, but also to centralize your most valuable insights for swift, coherent data analysis.
- You will improve your overall business efficiency by using the power of a robust data storage warehouse and working with high-power platforms that allow you to access all your





business-related insights from one intuitive centralized location, while quashing the roadblock of poor data performance.

# 11 CONCLUSION

The report analyzed data in the dashboards based on the set criteria. All three panels through user reviews received good feedback and there are no complaints about the information in them. Legal issues are also brought along with solutions to fix them.





# 12 REFERENCES

- Computerhope.com. (2020). Jupyter. https://www.computerhope.com/jargon/j/jupyter.htm
- Cordell, S. (2019). The Four Stages of Business Intelligence. https://blog.diamondit.com.au/types-of-bi
- Durcevic, S. (2019). *Common Business Intelligence Challenges Facing Entrepreneurs*. https://www.datapine.com/blog/business-intelligence-challenges/
- Eduardo Alonso. (2019). *Data Cleaning: What It Is and Why Your CRM Needs It*. https://www.vainu.com/blog/data-cleaning/#:~:text=Data%20cleaning%20or%20cleansing%20is,from%20a%20set%20of%20data.
- Expert System Team. (2020). What is Machine Learning? A Definition. https://www.expert.ai/blog/machine-learning-definition/
- Formplus Blog. (2020). Data Cleaning: Definition, Methods, and Uses in Research. https://www.formpl.us/blog/data-cleaning
- KnowledgeHut. (2019). *Pros & Cons Of Using Tableau For Your Projects*. https://www.knowledgehut.com/blog/business-intelligence-and-visualization/tableau-advantages-disadvantages
- Margaret Rouse. (2019). Data Labeling. https://whatis.techtarget.com/definition/data-labeling
- Melanie Chan. (2019). *The Benefits and Limitations of a Business Intelligence Dashboard*. https://www.unleashedsoftware.com/blog/benefits-limitations-business-intelligence-dashboard
- Pratt, M. K., & Fruhlinger, J. (2019). What is business intelligence? Transforming data into business insights. https://www.cio.com/article/2439504/business-intelligence-definition-and-solutions.html
- Rajesh Sampathkumar. (2019). What are the pros and cons of using Jupyter Notebook in Python? What did you like and what is didn't you like when you used it for analysis? https://www.quora.com/Whatare-the-pros-and-cons-of-using-Jupyter-Notebook-in-Python-What-did-you-like-and-what-is-didn-t-you-like-when-you-used-it-for-analysis
- Sandra Durcevic. (2020). *Get The Most Out Of Smart Business Intelligence Reporting* https://www.datapine.com/blog/business-intelligence-reporting/
- Sunil Ray. (2016). 7 Regression Techniques you should know! https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/#:~:text=Regression%20analysis%20is%20a%20form,effect%20relationship%20betwe en%20the%20variables.
- tableau.com. (2020). What you need to know about BI dashboards. https://www.tableau.com/learn/articles/business-intelligence/bi-dashboards#
- techterm.com. (2010). Python. https://techterms.com/definition/python
- techVidan.com. (2020). *Python Advantages and Disadvantages Step in the right direction*. https://techvidvan.com/tutorials/python-advantages-and-disadvantages/