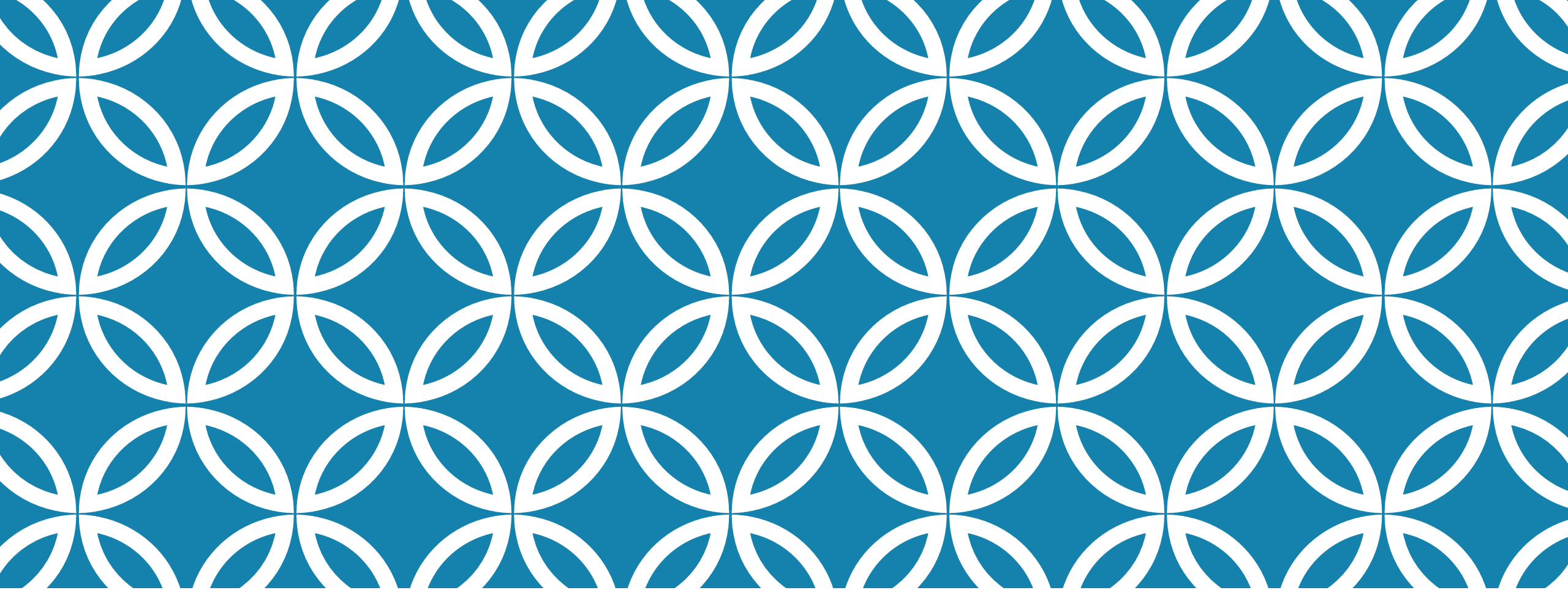


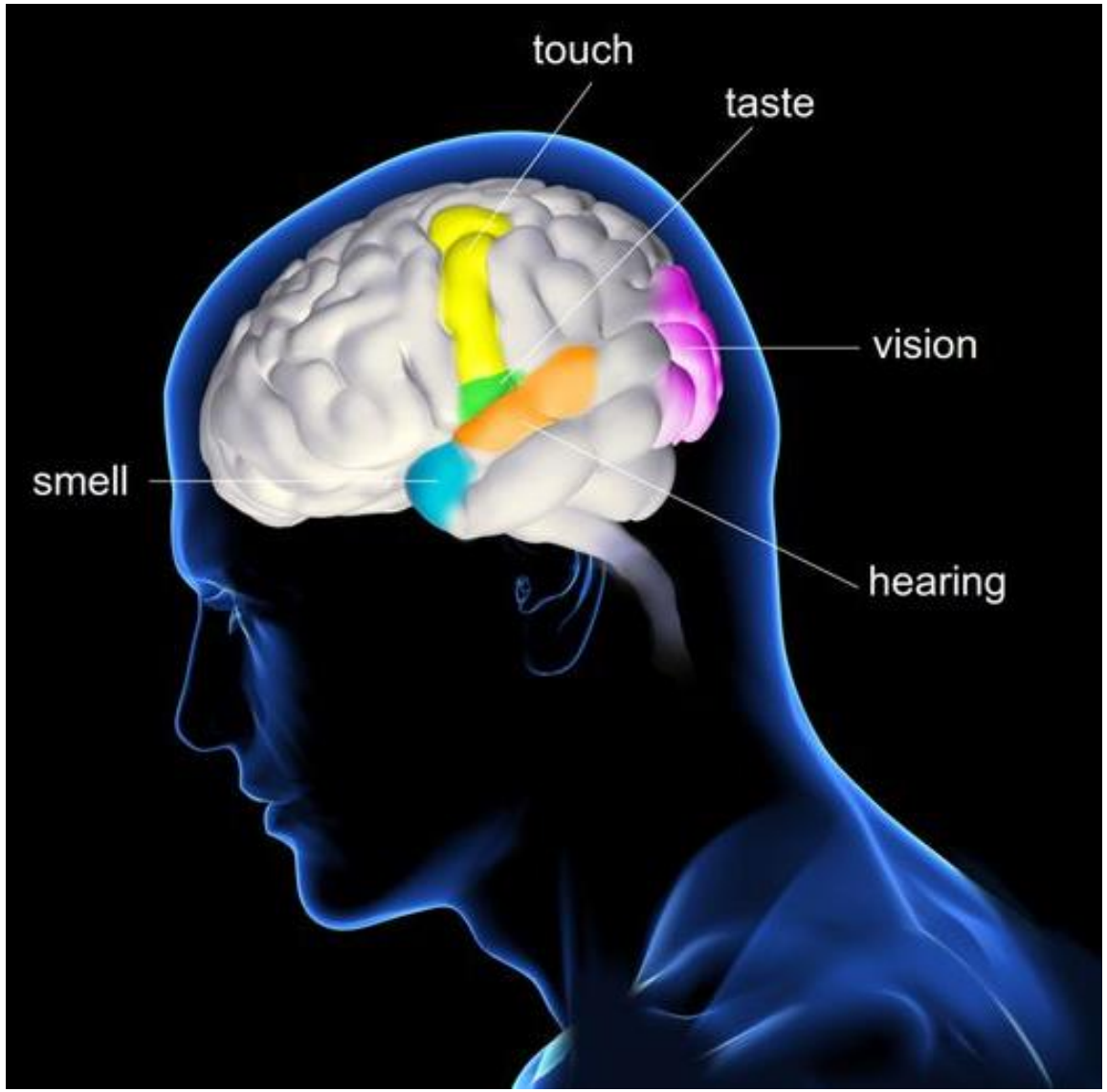
Data Introduction

WHAT IS DATA – INFORMAL DEFINITION

Human beings are exposed to data from birth.

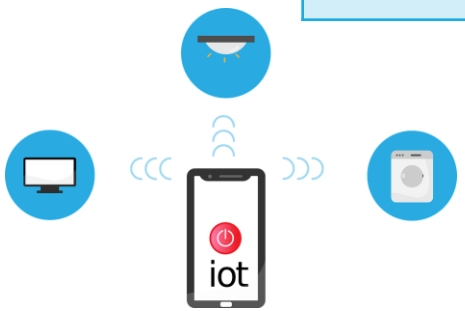
The eyes, ears, nose, skin, and tongue are continuously gathering various forms of data which the brain translates to sight, sound, smell, touch, and taste.

The brain then processes various forms of raw data it receives through sensory organs and translates it to speech, which is used to express opinion about the nature of raw data received.



DATA - ITS EVERYWHERE!!!

A collection of facts, such as numbers, words, measurements, observations or even just descriptions of things.



Bank & Finance		Astronomy	
	Healthcare		Surveys
Geology		Pharma	
	Bio Technology		IT & Telecom
Tourism	Iot		
		E Commerce	
Genetics	Manufacturing		Experiments
Airlines	Social Media	Sports	Movies

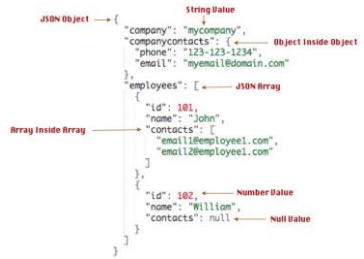
TYPES OF DATA

Structured Data can be displayed in Rows, Columns and Relational Databases. (Numbers, Dates & Strings)

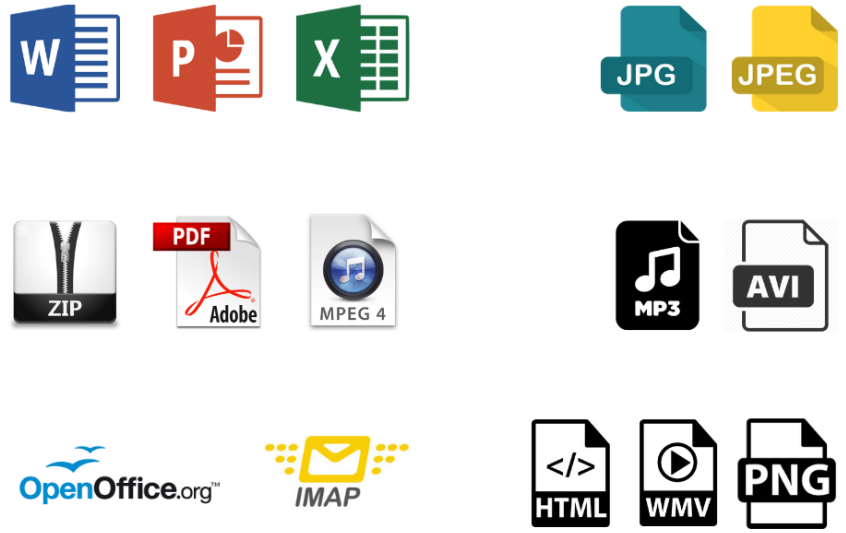
STRUCTURED

Name	Age	Salary
Jack	31	$2000
Jill	32	$1900
Jake	25	$1800

SEMI STRUCTURED



UNSTRUCTURED



STRUCTURED DATA

Columns/Variables
Descriptor Portion

Column1	Column2	Column3	Column4
A	01Jan2019	21	$2300
B	15Feb2019	32	$1900
C	01Aug2019	45	$1800
D	01Jun2019	12	$1200

Row/
Observations
Data Portion

Record/Data Point

Advantage of structured data is, its labelling to describe its attributes and relationships with other data.

Finance Table

Emp_Id	Position	Salary
001	SSE	$1200
002	MRG	$3000
003	SE	$900
004	SE	$1020

HR Table

Emp_Id	Name	Age	DoJ
001	Jack	31	23Jan2017
002	Jill	32	15Feb2019
003	Jake	25	0Aug2019
004	Sully	21	01Jun2019

Relation

Emp_Id	Name	Age	Dob
001	Jack	31	01Jan1989
002	Jill	32	15Feb1985
003	Jake	25	01Aug1990
004	Sully	21	01Jun1995

Fact Employee Table

Attendance Table

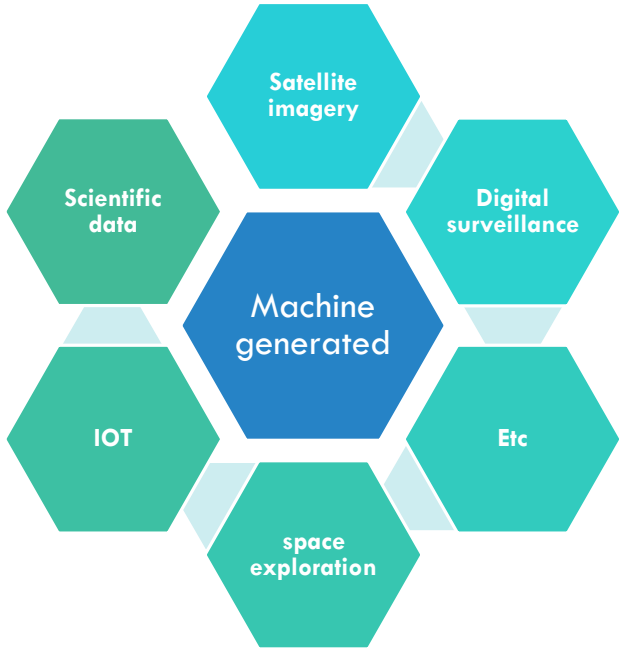
Emp_Id	Leave Taken	Swipe Miss Days
001	2	0
002	5	0
003	3	0
004	4	1

UNSTRUCTURED DATA

It has **an internal structure** (i.e. bits and bytes)

but is **not structured via pre-defined data models or schema**, i.e. not organized and labelled to identify meaningful relationships between data

It may be textual / non-textual. It may be human / machine-generated. It might also be stored within a non-relational database like NoSQL.



HOW MUCH DATA DO WE HAVE?

Social Media: (Every Minute)

- Snapchat users share 527,760 photos
- More than 120 professionals join LinkedIn
- Users watch 4,146,600 YouTube videos
- 456,000 tweets are sent on Twitter
- Instagram users post 46,740 photos

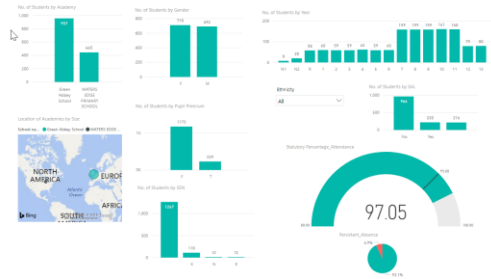
Internet:

- More than 3.7 billion humans use the internet.
- On average, Google now processes more than 40,000 searches EVERY second (3.5 billion searches per day)!
- Worldwide there are 5 billion searches a day.

DATA LIFECYCLE

Data Creation → ETL → Storage → BI → Analytics

Reports Historic Data. But, Fails to answer WHY Question!



Using Historic Data. Predictions are made!

DATA LIFE CYCLE

1. Data Creation:

➢Structured and Unstructured Data is created by many sources. Mostly, they are not human readable.

2. Extraction Transformation Loading:

➢ETL process converts the non-readable data into readable data and storing it in various databases.

3. Business Intelligence:

➢After ETL process, Business Intelligence systems provide historical, current, and predictive views of business operations, most often using data that has been gathered into a data warehouse or a data mart and occasionally working from operational data. Software elements support reporting, interactive “slice-and-dice” pivot-table analyses, visualization, and statistical data mining.

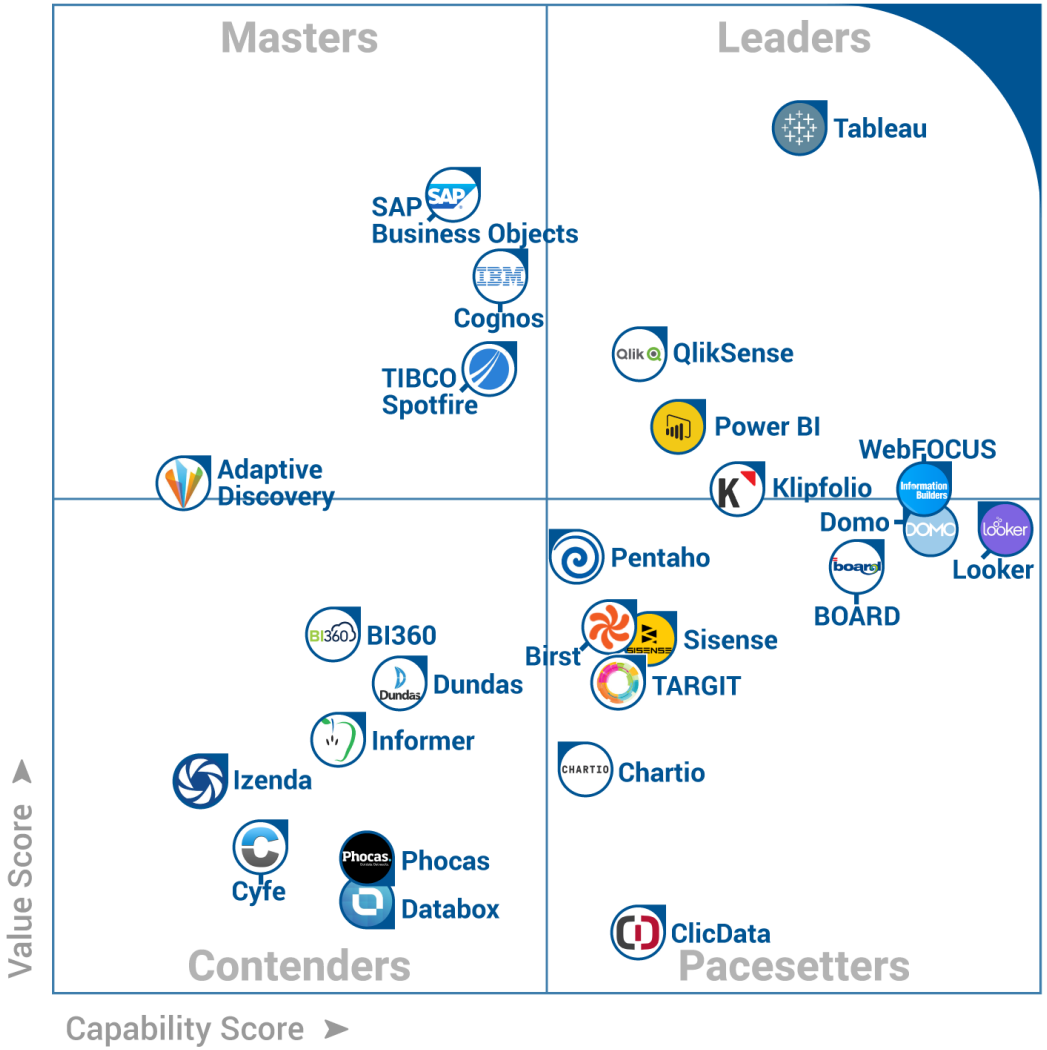
4. Data Analysis:

➢After ETL process, Data analysis is the process of evaluating data using analytical and statistical tools to discover useful information and aid in business decision making. There are a several data analysis methods including data mining, text analytics and data visualization.

ETL TOOLS IN MARKET

Oracle Warehouse Builder (OWB)
SAP Data Services
IBM InfosphereInformation Server
SAS Data Management
PowerCenterInformatica
Elixir Repertoire for Data ETL
Data Migrator (IBI)
SQL Server Integration Services (SSIS)
TalendStudio for Data Integration
SagentData Flow
ActianDataConnect
Open Text Integration Center
Oracle Data Integrator (ODI)
CognosData Manager
CloverETL
CenterpriseData Integrator
IBM InfosphereWarehouse Edition
Pentaho Data Integration
AdeptiaIntegration Server
QlikView Expressor

BUSINESS INTELLIGENCE TOOLS IN MARKET



ANALYTICS TOOLS IN MARKET

WHAT TO DO WITH THESE DATA?

WHY IS DATA ANALYSIS IMPORTANT?

- Companies makes attempts to gather data, for instance, by monitoring its competitors' performance, sales figures, and buying trends etc. in an effort to be more competitive. However, nobody can understand customers' behaviors and competitors' performance without the skills to analyze all that data.
- Data analysis, therefore, is a necessity for making well-informed and efficient decisions.
- Data analysis is what helps organizations determine their positions in the market relative to competitors.
- It is what helps us identify the potential risks that need to be avoided and the opportunities that must be grabbed in order to grow.
- It is, in fact, data analysis that enables us to gauge the satisfaction level of the customers and their needs in order to come up with new products and services that provide greater satisfaction to them. Therefore, it is an understatement to say that data analysis is important for the success of businesses.

Thank You!