INT233 PROJECT REPORT

(Project Semester August-December 2020)

PROJECT MANAGEMENT

Submitted by

Mohammed Kareem

11911638

Programme and Section-EM152

Course Code– INT233

Under the Guidance of

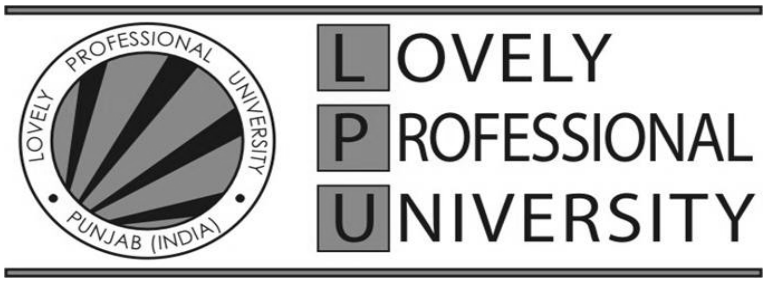
Richa Sharma : 27957

Assistant Professor

Discipline of SEEE

Lovely School of Electronics and Electrical Engineering

Lovely Professional University, Phagwara



CERTIFICATE

This is to certify that Mohammed Kareem bearing Registration no.11911638 has completed INT233 project titled, **Project management** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Richa Sharma : 27957
Assistant Professor
School of Electronics and Electrical Engineering
Lovely Professional University
Phagwara, Punjab.

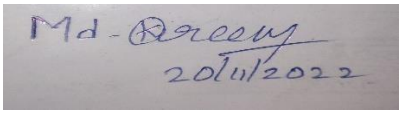
Date: 20-11-2022

DECLARATION

I, Mohammed Kareem student of Electronics and Electrical Engineering under SEEE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 21-11-2022

Signature



Registration No-11911638 Name of the student Mohammed Kareem

TABLE OF CONTENT :

INTRODUCTION

DATA

Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things. Everyday a huge amount of data is being produced. It needs to be stored and analyzed to draw some insights from it and that is helpful in optimizing the performance in Businesses.

The data is classified into majorly four categories:

- Nominal data.
- Ordinal data.
- Discrete data.
- Continuous data.

Qualitative vs Quantitative

Data can be qualitative or quantitative.

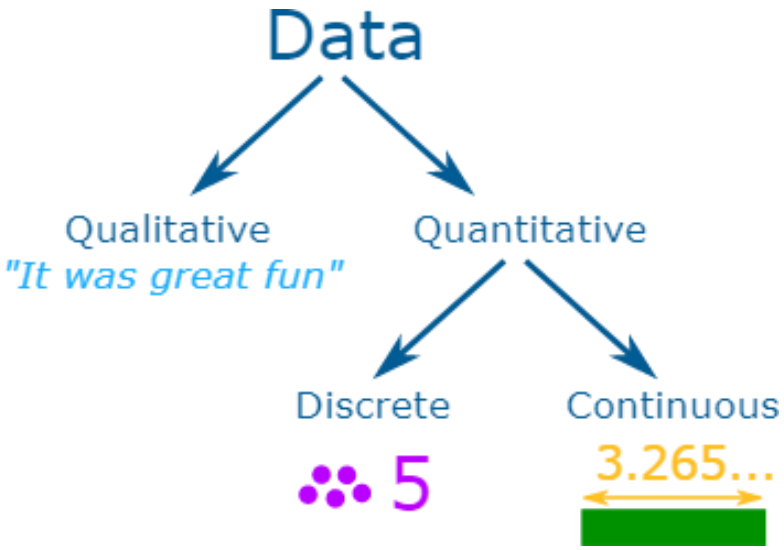
- Qualitative data is descriptive information (it describes something)
- Quantitative data is numerical information (numbers)

Quantitative data can be Discrete or Continuous:

Discrete data can only take certain values (like whole numbers)

Continuous data can take any value (within a range)

Discrete data is counted, Continuous data is measured



Example: What do we know about Arrow the Dog?

Qualitative:

- He is brown and black
- He has long hair
- He has lots of energy

Quantitative:

Discrete:

- He has 4 legs
- He has 2 brothers

Continuous:

- He weighs 25.5 kg
- He is 565 mm tall

A datum is an individual value in a collection of data. Data are usually organized into structures such as tables that provide additional context and meaning, and which may themselves be used as data in larger structures. Data may be used as variables in a computational process. Data may represent abstract ideas or concrete measurements. Data are commonly used in scientific research, finance, and in virtually every other form of human organizational activity. Examples of data sets include stock prices, crime rates, unemployment rates, literacy rates, and census data. In this context, data represents the raw facts and figures which can be used in such a manner in order to capture the useful information out of it.

Data are collected using techniques such as measurement, observation, query, or analysis, and typically represented as numbers or characters which may be further processed. Field data are data that are collected in an uncontrolled in-situ environment. Experimental data are data that are generated in the course of a controlled scientific experiment. Data is analyzed using techniques such as

calculation, reasoning, discussion, presentation, visualization, or other forms of post-analysis. Prior to analysis, raw data (or unprocessed data) is typically cleaned: Outliers are removed and obvious instrument or data entry errors are corrected.

DATA ANALYSIS

What is data analysis

Data analysis is the process of cleaning, changing, and processing raw data and extracting actionable, relevant information that helps businesses make informed decisions. The procedure helps reduce the risks inherent in decision-making by providing useful insights and statistics, often presented in charts, images, tables, and graphs.

A simple example of data analysis can be seen whenever we make a decision in our daily lives by evaluating what has happened in the past or what will happen if we make that decision. Basically, this is the process of analyzing the past or future and making a decision based on that analysis.

Why is Data Analysis Important?

list of reasons why data analysis is crucial to doing business today.

- Better Customer Targeting
- Reduce Operational Costs
- Better Problem-Solving Methods
- You Get More Accurate Data

What Is the Data Analysis Process?

The process of data analysis, or alternately, data analysis steps, involves gathering all the information, processing it, exploring the data, and using it to find patterns and other insights. The process of data analysis consists of

Data Requirement Gathering

Data Collection:

Data Cleaning:

Data Analysis

Data Interpretation

Data Visualization

What is Data Visualization?

Data visualization is the process of communicating and translating data and information in a visual context, usually employing a graph, chart, bar, or other visual aid. Visualization also uses images to communicate the relationships between various sets of data.

Data visualization is also called information visualization, information graphics, and statistical graphics. It is a step in the process of data science, which tells us that after all data has been collected, processed, and modeled, the information must be visualized so that users can use it to draw conclusions.

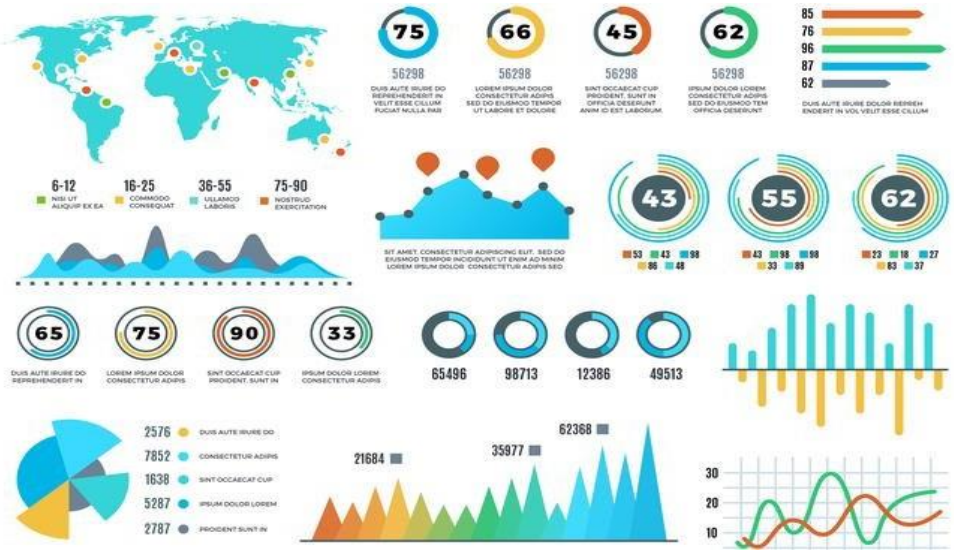
Also, data visualization is part of the broader discipline of data presentation architecture (DPA), whose purpose is to identify, find, manipulate, format, and deliver data in the best way possible.

The Importance of Data Visualization

Visually representing insights derived from data provides a means for people to see and understand data patterns, trends, and outliers. Data visualization draws us that picture, presenting facts and figures in a clear, visually appealing manner.

The benefits of data visualization include:

- Gives the reader the means to quickly absorb information, improve insights and make faster decisions
- Provides an easy means of distributing information that offers users more opportunities to share their insights with everyone involved in the project
- Imparts an increased understanding of what steps an organization must take to improve itself
- Offers the ability to attract and maintain the audience's interest by giving them the information they can understand
- Gives the decision-makers the means to quickly act on findings, deliver successful outcomes faster, and have fewer errors
- Eliminates the need to excessively rely on data scientists because it is more accessible and easily understood.

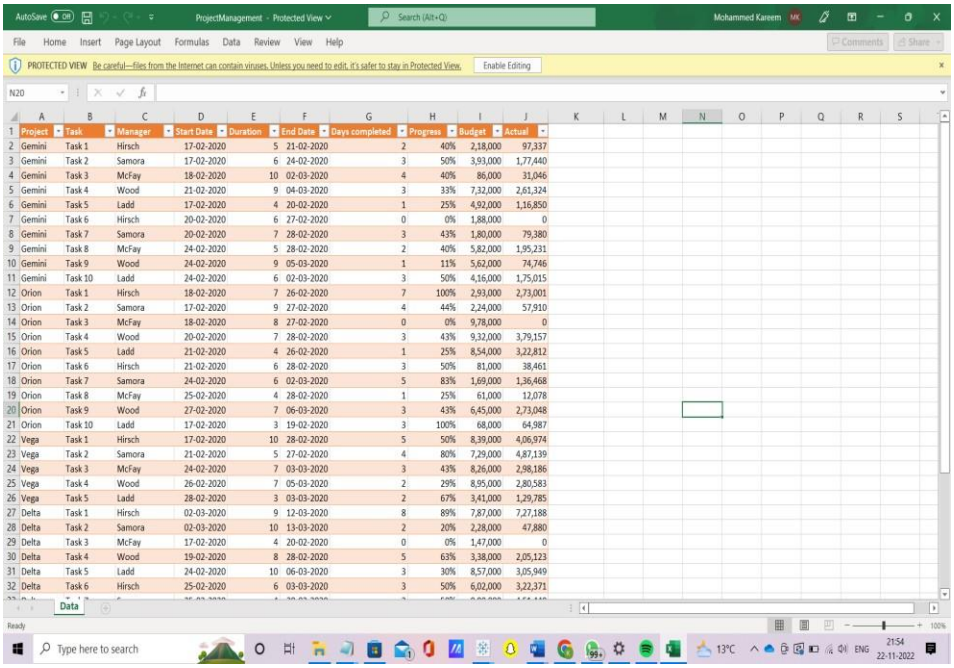


A sample data visualization

The dataset I used is Project management data set from Kaggle

Columns in this dataset are:

1. Project
2. Task
3. Manager
4. Start Date
5. Duration
6. End Date
7. Days completed
8. Progress
9. Budget
10. Actual



Project	Task	Manager	Start Date	Duration	End Date	Days completed	Progress	Budget	Actual
Gemini	Task 1	Hirsch	17-02-2020	5	21-02-2020	2	40%	2,18,000	97,337
Gemini	Task 2	Samora	17-02-2020	6	24-02-2020	3	50%	3,93,000	1,77,440
Gemini	Task 3	McFay	18-02-2020	10	02-03-2020	4	40%	86,000	31,046
Gemini	Task 4	Wood	21-02-2020	9	04-03-2020	3	33%	7,32,000	2,61,324
Gemini	Task 5	Ladd	17-02-2020	4	20-02-2020	1	25%	4,92,000	1,16,850
Gemini	Task 6	Hirsch	20-02-2020	6	27-02-2020	0	0%	1,88,000	0
Gemini	Task 7	Samora	20-02-2020	7	28-02-2020	3	43%	1,80,000	79,380
Gemini	Task 8	McFay	24-02-2020	5	28-02-2020	2	40%	5,82,000	1,95,231
Gemini	Task 9	Wood	24-02-2020	9	05-03-2020	1	11%	5,62,000	74,746
Gemini	Task 10	Ladd	24-02-2020	6	02-03-2020	3	50%	4,16,000	1,75,015
Orion	Task 1	Hirsch	18-02-2020	7	26-02-2020	7	100%	2,93,000	2,73,001
Orion	Task 2	Samora	17-02-2020	9	27-02-2020	4	44%	2,24,000	57,910
Orion	Task 3	McFay	18-02-2020	8	27-02-2020	0	0%	9,78,000	0
Orion	Task 4	Wood	20-02-2020	7	28-02-2020	3	43%	9,32,000	3,79,157
Orion	Task 5	Ladd	21-02-2020	4	26-02-2020	1	25%	8,54,000	3,22,812
Orion	Task 6	Hirsch	21-02-2020	6	28-02-2020	3	50%	81,000	38,461
Orion	Task 7	Samora	24-02-2020	6	02-03-2020	5	83%	1,69,000	1,36,468
Orion	Task 8	McFay	25-02-2020	4	28-02-2020	1	25%	61,000	12,078
Orion	Task 9	Wood	27-02-2020	7	06-03-2020	3	43%	6,45,000	2,73,048
Orion	Task 10	Ladd	17-02-2020	3	19-02-2020	3	100%	68,000	64,987
Vega	Task 1	Hirsch	17-02-2020	10	28-02-2020	5	50%	8,39,000	4,06,974
Vega	Task 2	Samora	21-02-2020	5	27-02-2020	4	80%	7,29,000	4,87,139
Vega	Task 3	McFay	24-02-2020	7	03-03-2020	3	43%	8,26,000	2,98,186
Vega	Task 4	Wood	26-02-2020	7	05-03-2020	2	29%	8,95,000	2,80,583
Vega	Task 5	Ladd	28-02-2020	3	03-03-2020	2	67%	3,41,000	1,29,785
Delta	Task 1	Hirsch	02-03-2020	9	12-03-2020	8	89%	7,87,000	7,27,188
Delta	Task 2	Samora	02-03-2020	10	13-03-2020	2	20%	2,28,000	47,880
Delta	Task 3	McFay	17-02-2020	4	20-02-2020	0	0%	1,47,000	0
Delta	Task 4	Wood	19-02-2020	8	28-02-2020	5	63%	3,38,000	2,05,123
Delta	Task 5	Ladd	24-02-2020	10	06-03-2020	3	30%	8,57,000	3,05,949
Delta	Task 6	Hirsch	25-02-2020	6	03-03-2020	3	50%	6,02,000	3,22,371

Preview of Project management Dataset

OBJECTIVES / SCOPE OF THE ANALYSIS:

1. Aim of this analysis is to give results of the following analysis and give future trends:
2. Budget for Project
3. Task for Project
4. Manager took total Days to complete that project
5. Progress of Tasks
6. Budget
7. Actual
8. Budget for task
9. Actual for task
10. Days took to complete task

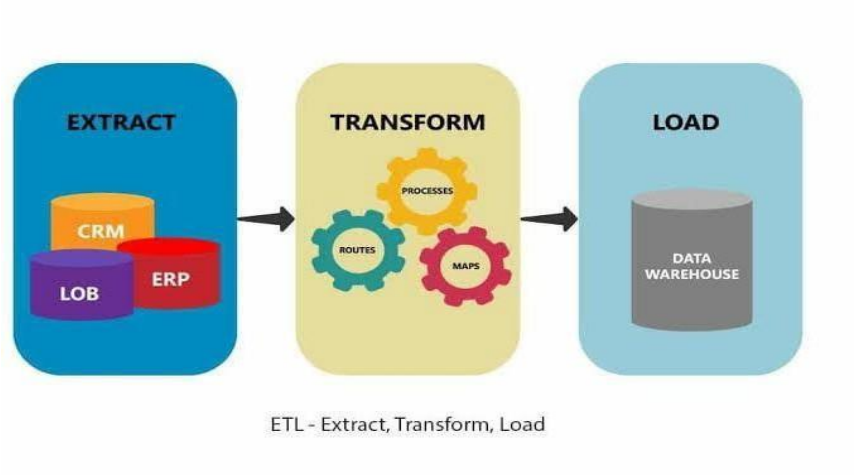
ETL PROCESS :

Extract, transform, and load (ETL) are 3 data processes, followed after data collection.

Extraction takes data, collected in data sources like flat files, databases (relational, hierarchical etc.), transactional datastores, semi-structured repositories (e.g. email systems or document libraries) with different structure and format, pre-validating extracted data and parsing valid data to destination (e.g. staging database)

Transformation takes extracted data and applies predefined rules and functions to it, including selection (e.g. ignore or remove NULLs), data cleansing, encoding (e.g. mapping "Male" to "M"), deriving (e.g. calculating designated value as a product of extracted value and predefined constant) , sorting, joining data from multiple sources (e.g. lookup or merge), aggregation (e.g. summary for each month), transposing (columns to rows or vice versa), splitting, disaggregation, lookups (e.g. validation through dictionaries), predefined validation etc. which may lead to rejection of some data. Transformed data can be stored into Data Warehouse (DW).

Load takes transformed data and places it into end target, in most cases called Data Mart (sometimes they called Data Warehouse too). Load can append, refresh or/and overwrite pre-existing data, apply constraints and execute appropriate triggers (to enforce data integrity, uniqueness, mandatory fields, provide log etc.) and may start additional processes, like databackup or replication.



ETL - Extract, Transform, Load

Tableau Prep

Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing your data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

Start by connecting to your data from a variety of files, servers, or Tableau extracts. Connect to and combine data from multiple data sources. Drag and drop or double-click to bring your tables into the flow pane, and then add flow steps where you can then use familiar operations such as filter, split, rename, pivot, join, union and more to clean and shape your data.

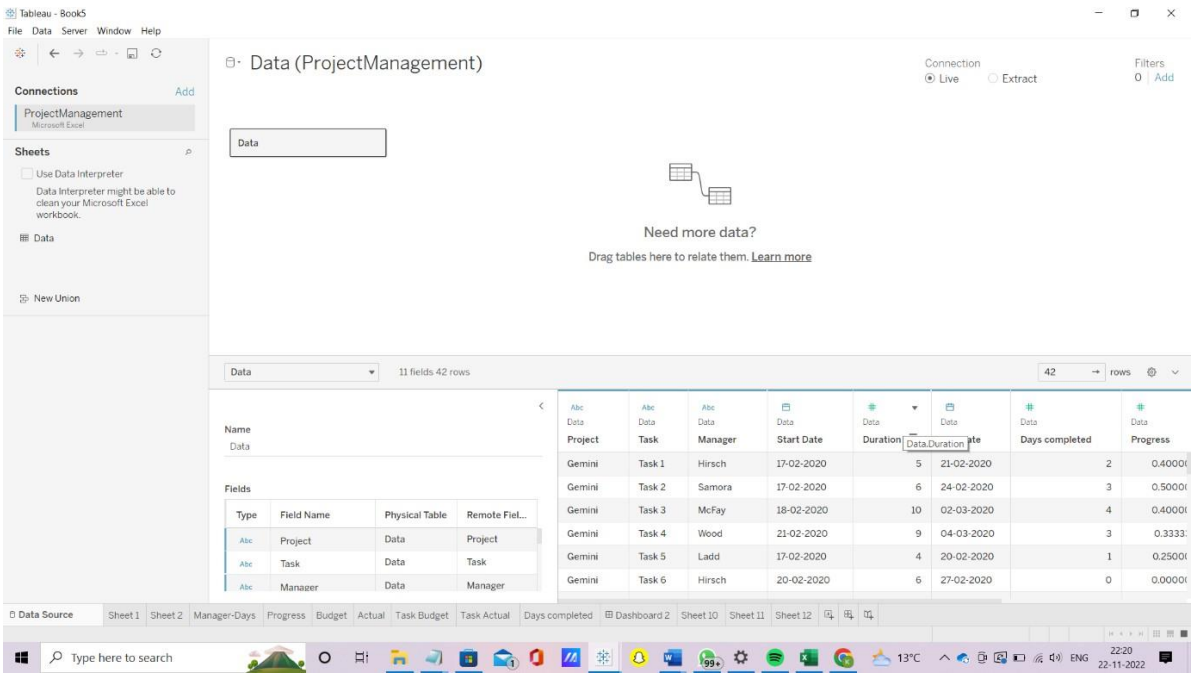
How Does Tableau Prep Work?

Tableau Prep helps you **examine and visualize your data**, enabling you to do the following:

1. **Connect and extract data**
2. **Understand data:**
    a. Number of columns/fields in your data
    b. Number of records
    c. Data types of fields
    d. Number of distinct values in a field
    e. Visualize how the data is distributed by field
3. **Identify issues and errors**
4. **Clean/Modify and Filter data**
    a. Rename fields
    b. Remove fields
    c. Modify/change values in a field
    d. Split fields
    e. Aggregate data
    f. Filter out data

5. **Enhance data**
    a. Add Calculated fields
    b. Join additional data
    c. Union additional data
6. **Output resulting data for use in analysis and reporting**

And, once you get new data (as long as it is in the same format and same field names), the ETL process you created is reusable. No longer will you have to repeat the process and steps necessary to transform your data each time the source data is updated, instead the ETL process flow has all the steps and logic you built. All you need to do is re-run the flow to get the new data output, resulting in many hours saved from data processing and cleansing, which can be used for analysis instead!



ANALYSIS ON DATASET:

1.MANAGER DAYS

i. Introduction :

This visualization shows us the manager took no.of days for that project .

ii. General Description :

Here I've dropped sum(Budget) in columns and project in rows.

I used pie chart show results.

iii. Specific Requirements, functions and formulas and prediction models :

I customized **project** in colour to show the result in more readable way.

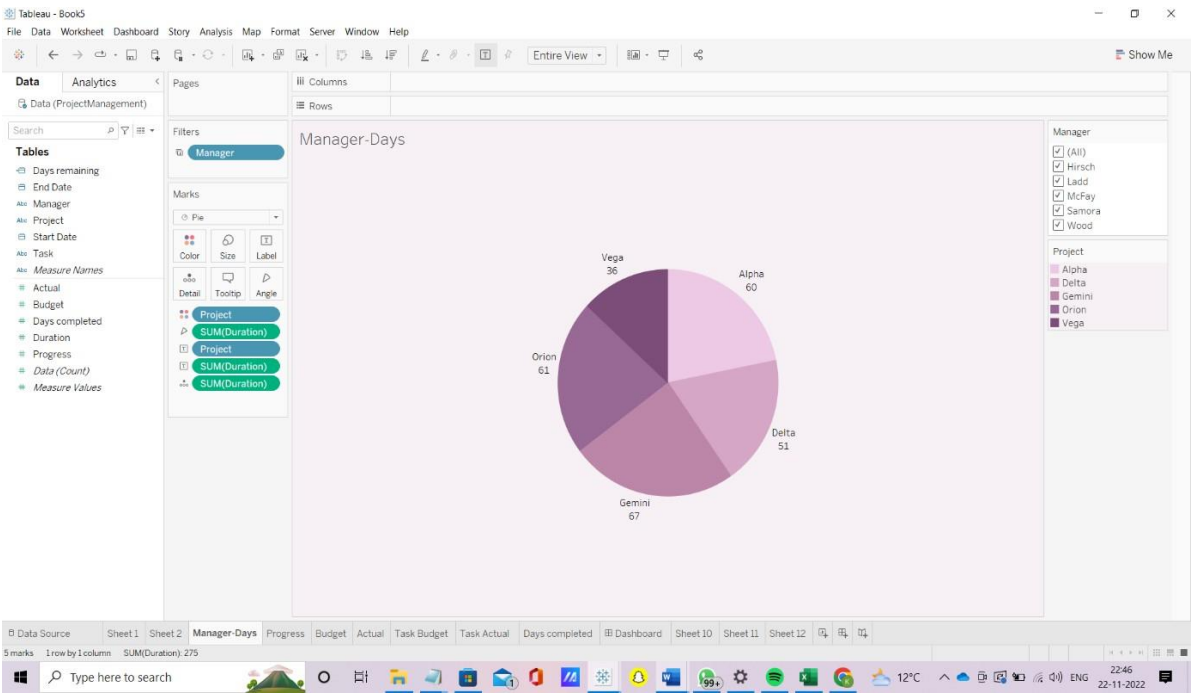
Then duration in text, detail and in angle.

Added manager in filters.

iv. Analysis results :

By this we can easily get the total days took by the particular manager

v. Visualization :



Progress

I. Introduction :

This visualization shows us the progress of the manager for that particular project

II. General Description :

I have dropped budget in marks of color and size.

I have used tree map to show the visualization

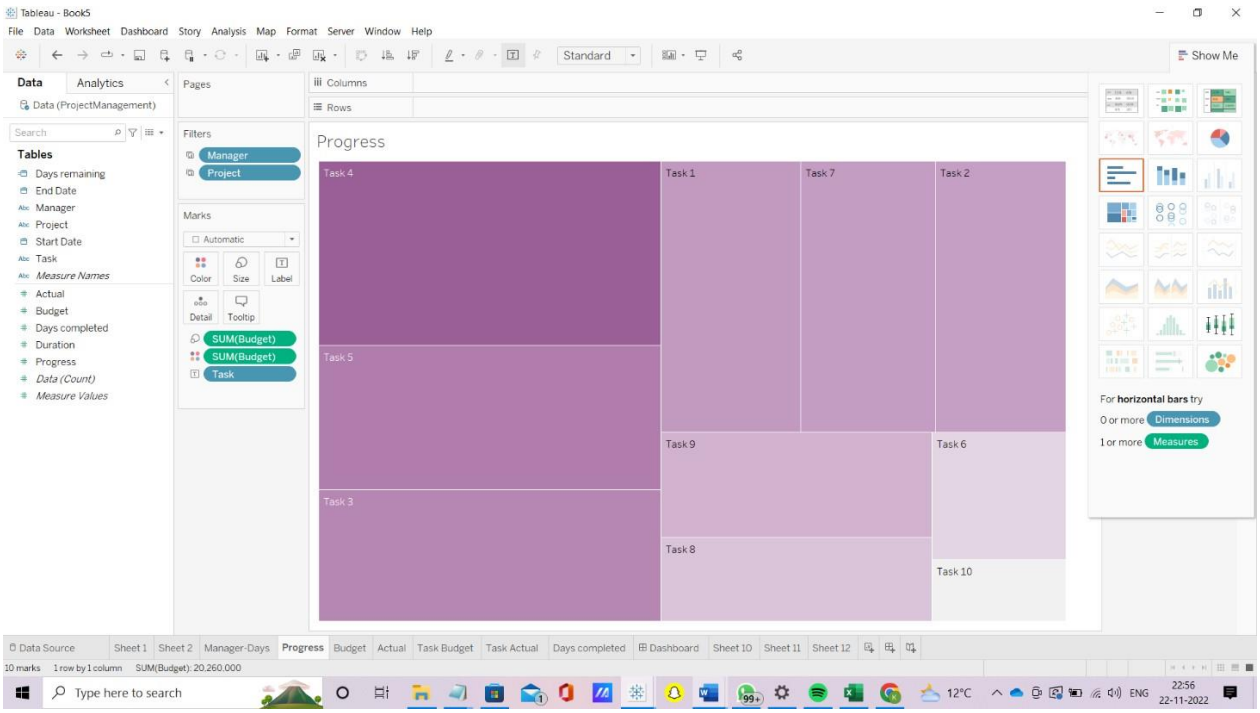
III. Specific Requirements, functions and formulas and prediction models :

I used manager and project in filters

IV. Analysis results :

In this the Task that has highest budget is shown in darker color.

V. Visualization :



Budget

i. Introduction :

This visualization shows the budget took by the manager and project

ii. General Description :

I have dropped budget in columns

I used bar chart to show this information

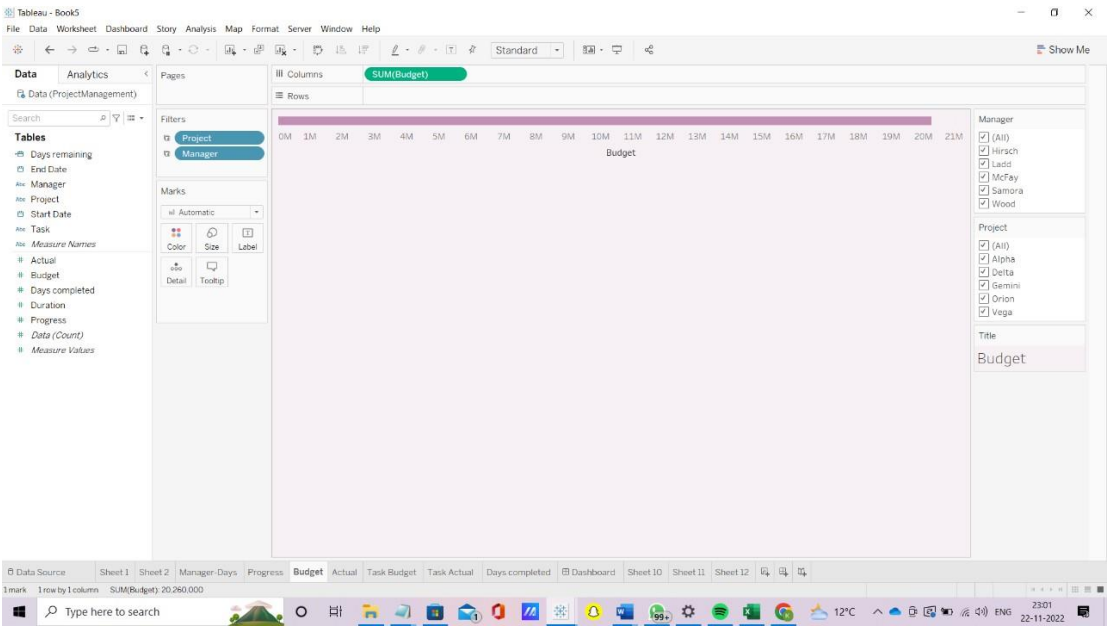
iii. Specific Requirements, functions and formulas and prediction models :

I dropped project and manager in filters

iv. Analysis results :

by this we can easily know the budget took for the project and by the manager

v. Visualization :



Budget

i. Introduction :

This visualization shows the actual took by the manager and project

ii. General Description :

I have dropped actual in columns

I used bar chart to show this information

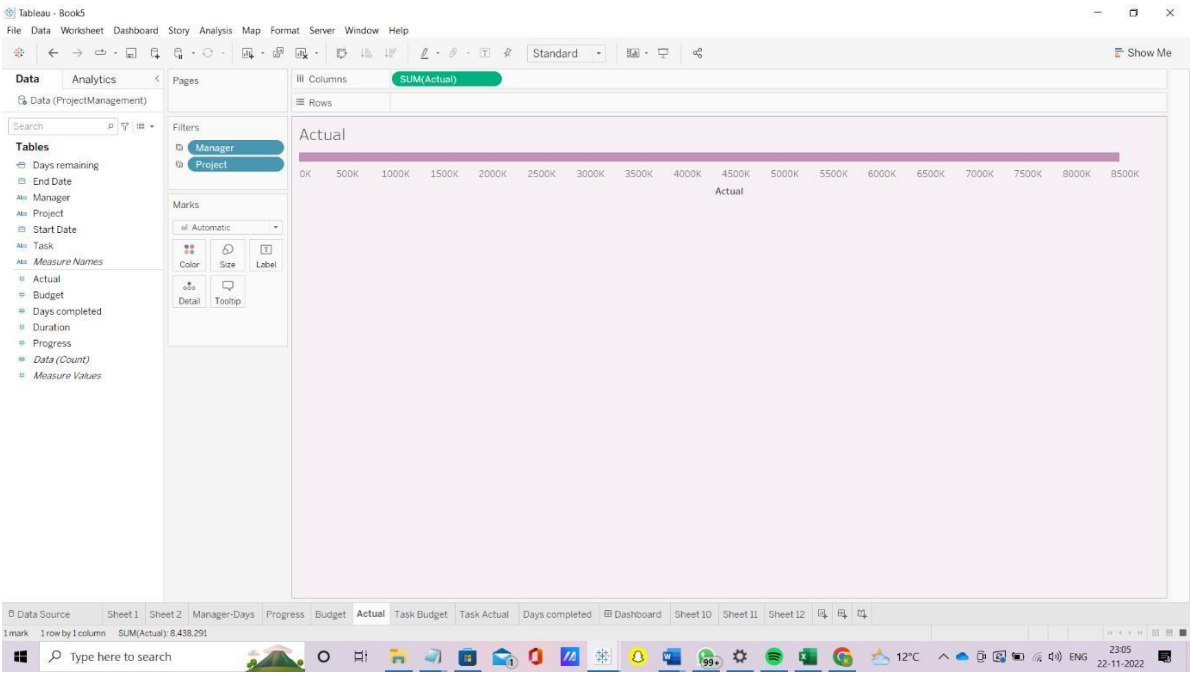
iii. Specific Requirements, functions and formulas and prediction models :

I dropped project and manager in filters

iv. Analysis results :

by this we can easily know the actual took for the project and by the manager

v. Visualization :



I. Introduction:

This visualization shows the budget took by the task

II. General Description :

I dropped budget in the rows

I used bar chart to show this visualization

III. Specific Requirements, functions and formulas and prediction models :

Added task in color, budget in the size and text

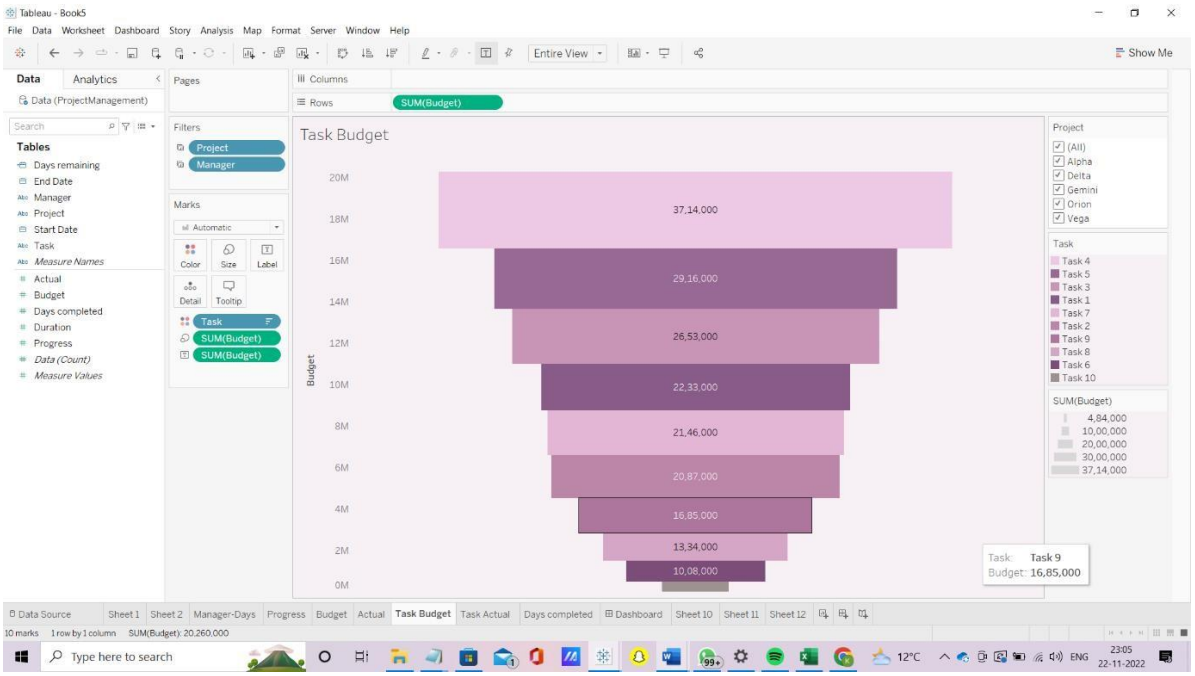
Then arrange it in descending order

Added project and manager in filter

IV. Analysis results :

This looks like funnel that shows the budget took for the particular task

V. Visualization :



Introduction:

This visualization shows the budget took by the task

I. General Description :

I dropped Actual in the rows

I used bar chart to show this visualization

II. Specific Requirements, functions and formulas and prediction models :

Added task in color, Actual in the size and text

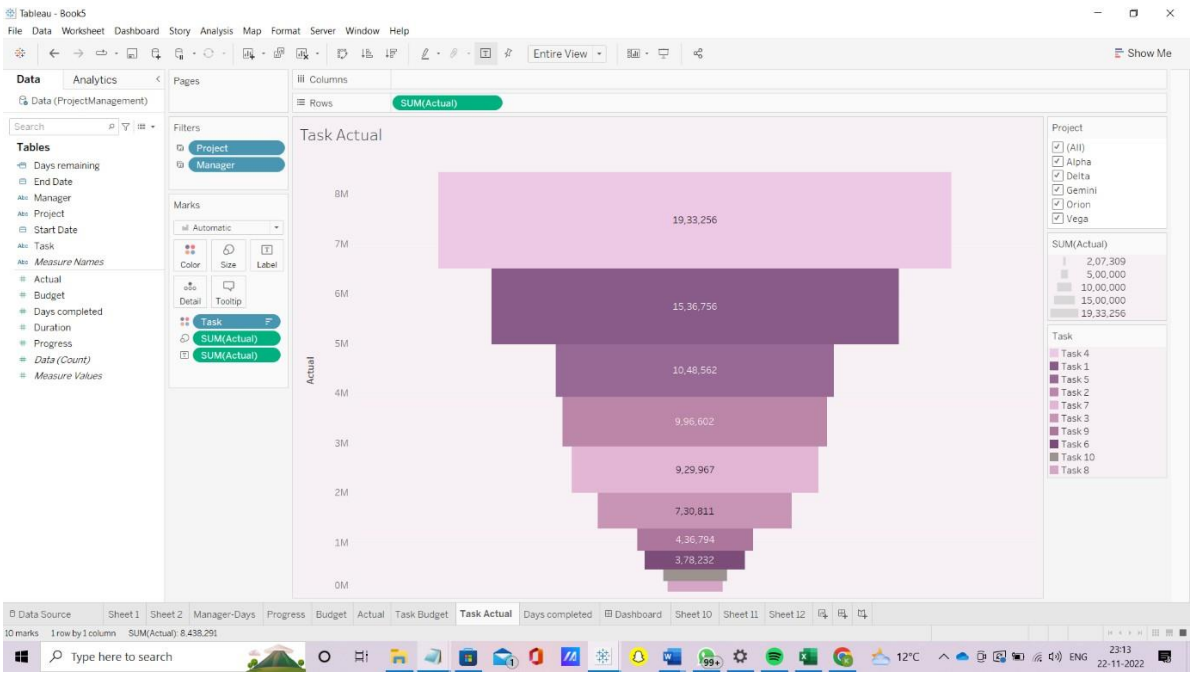
Then arrange it in descending order

Added project and manager in filter

III. Analysis results :

This looks like funnel that shows the actual took for the particular task

IV. Visualization :



Days Completed

I. Introduction :

This is the visualization of days took to complete task

II. General Description :

I dropped task in columns and rows in days completed

I used bar chart to show this visualization.

III. Specific Requirements, functions and formulas and prediction models :

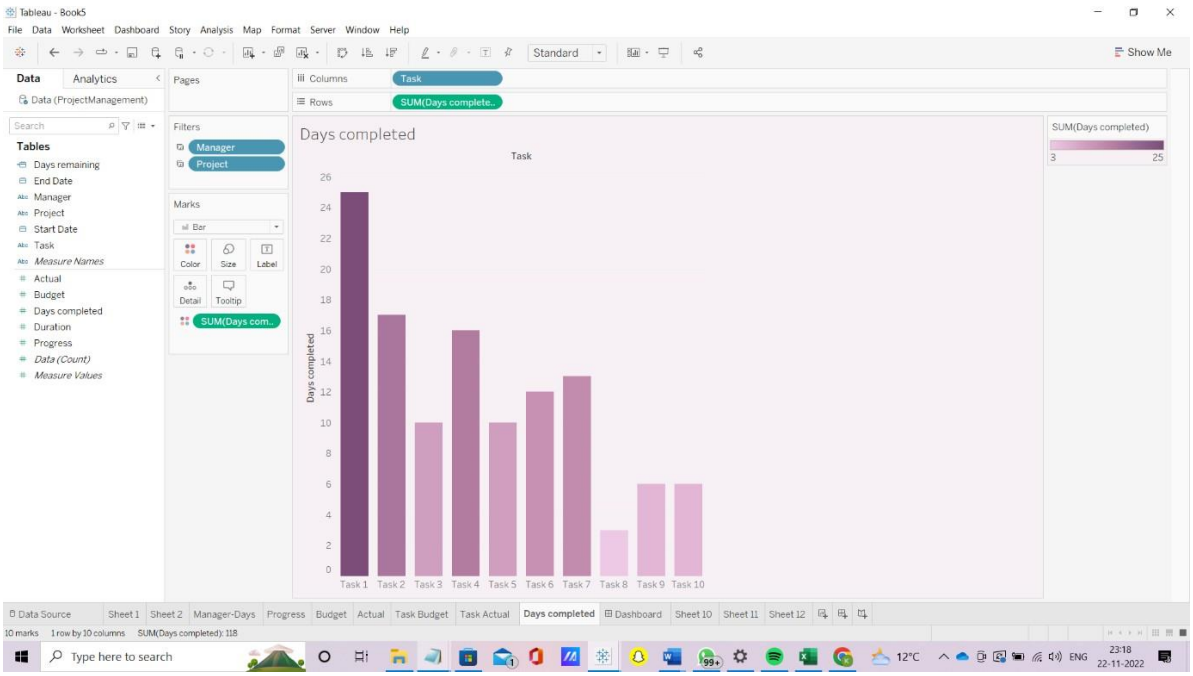
I added days completed in marks of color

I used manager and project in filters

IV. Analysis results :

By this we can easily task took no.of days to get completed by the manager and project.

V. Visualization :



DASHBOARD

The final dash looks like this

LIST OF ANALYSIS WITH RESULTS :

S.No	ANALYSIS	RESULT
1	Manager days	Manager wood took highest no of days 24 to complete project alpha
2	Project	Task 4 took highest budget of 36,14,000
3	Budget	The total budget is 2,02,60,000
4	Actual	The total actual is 84,38,291
5	Task-budget	Task 4 took highest budget of 37,14,000
6	Task actual	Task 4 took highest actual of 19,33,256
7	Days completed	Task 1 took more to complete i.e 25 days.

REFERNCES :

Kaggle dataset

BIBLIOGRPHY :

Referred following websites :

https://help.tableau.com/

https://www.analyticsvidhya.com/

https://www.tutorialspoint.com/

https://www.mathsisfun.com/data/data.html