**PLACEMENT DATA ANALYSIS**

|  |  |
| --- | --- |
|  | **2024** |
|  | **Internship Project**    **MBA BA I Year II Trimester** |

The project Placement Data Analysis performed on Anurag University MBA department placement data aims to develop an integrated data analysis framework to optimize university placement operations and enhance student experiences through data-driven insights, striving to improve student placement rates. Leveraging data analytics techniques, the project explores various dimensions such as MBA specializations, ICET ranks, gender disparities, and company placements. By examining placement trends, identifying influential factors, and assessing student attributes, the project seeks to optimize placement strategies and improve the effectiveness of placement initiatives.

**Contents**

Project cover sheet.................................................................................................... 2

Executive Summary ................................................................................................. 3

Objective .................................................................................................................. 3

Project Milestones .................................................................................................. 25

Requirements Specifications .................................................................................. 28

Business Process Chart ........................................................................................... 32

Activities ................................................................................................................ 35

Data Flows .............................................................................................................. 37

Data Model (ER) .................................................................................................... 40

Datasets ................................................................................................................... 45

Data Points: Dimensions & Measures .................................................................... 52

ETL ......................................................................................................................... 53

Visualizations ......................................................................................................... 55

Visualization Dashboards, Charts and graphs : .................................................. 55

# Project cover sheet

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Title** | Analytics of Academic Excellence | **Course** | MBA BA I Year – II Trimester |
| **Start Date** | 19-02-2024 | **Guide/Client** | Dr Bheemiah |
| **End Date** | 26-04-2024 | **Team Lead** | Anvitha |

**Team Members**

|  |  |
| --- | --- |
| **SNO** | **Name** |
| 1. | A. Chandra paul |
| 2. | G. Anvitha |
| 3. | K. Ajitha Reddy |
| 4. | Khaja Sami Uddin |
| 5. | K. Vishal |
| 6. | N. Vishnu Vardhan |

Signature:

Guide/ClientName:

Date:

# Executive Summary

* Explore placement outcomes based on MBA specialization (HR, Finance, Marketing ).
* Identifying which specialization has the maximum number of placements.
* Analyze the relationship between ICET rank and student placements.
* Investigate gender disparities in placement outcomes.
* Determine the number of offers received by students and their distribution across various attributes.
* Analyze placement patterns pertaining to companies.

# Objective

The objectives of the university data analytics project encompass a comprehensive exploration of placement outcomes and factors influencing them. Here's how each objective contributes to the overall project:

Explore placement outcomes based on MBA specialization (HR, Finance, Marketing):

This objective involves analyzing placement rates, average salaries, and industry preferences for each MBA specialization.

Insights gained help understand the demand for different skill sets and tailor placement strategies accordingly.

Identifying which specialization has the maximum number of placements:

By determining the specialization with the highest number of placements, the university can identify areas of strength and potential opportunities for growth.

This insight informs resource allocation and curriculum development to meet industry demands.

Analyze the relationship between ICET rank and student placements:

Understanding the correlation between ICET rank and placements provides insights into the role of academic performance in securing placements.

It helps identify trends and patterns that influence placement outcomes and guide academic support programs.

Investigate gender disparities in placement outcomes:

Analyzing gender disparities in placements highlights any inequities or biases in the hiring process.

It enables the university to implement targeted interventions to support underrepresented groups and promote inclusivity.

Determine the number of offers received by students and their distribution across various attributes:

By analyzing the distribution of offers across different attributes such as specialization, gender, and academic performance, the university can identify trends and preferences among recruiters.

Insights inform strategies to improve students' competitiveness in the job market.

Analyze placement patterns pertaining to companies:

This objective involves examining placement trends, preferences, and relationships with companies.

Insights help strengthen partnerships with high-demand recruiters, diversify placement opportunities, and enhance industry connections.

### • Analyzing real world data

Analyzing real-world data involves extracting insights from raw, unstructured information collected from various sources, such as databases, sensors, or social media. It holds immense importance as it enables evidence-based decision-making, identifies trends, patterns, and anomalies, and informs strategies across industries, from healthcare to finance. By analyzing real-world data, organizations can optimize operations, improve customer experiences, and innovate products or services. It facilitates understanding of complex systems, predicting future outcomes, and mitigating risks. Real-world data analysis is essential for staying competitive in today's data-driven world and fostering innovation and growth.

Analyzing real-world data in the university data analytics project is crucial as it provides insights into actual placement trends, student preferences, and industry dynamics. By examining real data, we can uncover patterns, correlations, and factors influencing placement outcomes, enabling informed decision-making and strategy development. Real-world data analysis helps identify areas for improvement, optimize placement strategies, and tailor interventions to meet the diverse needs of students and recruiters. It enhances the project's relevance by grounding it in practical insights derived from authentic experiences and outcomes. Analyzing realworld data empowers stakeholders to make data-driven decisions, foster student success, and enhance the effectiveness of placement programs.

### • Developing Data Models

Data modeling is vital as it structures raw data into organized formats, facilitating analysis, interpretation, and decision-making. Key aspects include defining entities, attributes, and relationships to represent real-world phenomena accurately. It enables effective data integration, aggregation, and visualization, enhancing understanding and insights generation. Well-designed data models ensure data consistency, integrity, and scalability, supporting efficient storage and retrieval of information. Ultimately, data modeling lays the foundation for building robust analytical solutions, driving innovation, and achieving business objectives. In the context of the university data analytics project, developing data models involves creating structured representation.

Model Designing:

Entity-Relationship (ER) Modeling: Create an ER diagram to visually represent the relationships between entities and their attributes.

Dimensional Modeling: Design dimensional models such as star schemas or snowflake schemas to organize data into dimensions (e.g., time, student) and measures (e.g., number of placements, average salary).

Hierarchical Modeling: Develop hierarchical models to represent data hierarchies, such as student education levels or company hierarchies.

### • Identifying Dimensions and Measures

The dimensions and measures provide a structured framework for analyzing and deriving insights from the placement data. By exploring the relationships between dimensions and measures, universities can gain a comprehensive understanding of placement trends, identify areas for improvement, and optimize their placement strategies effectively.

Dimensions: Year of Placement: This dimension categorizes the data based on the year in which placements occurred (e.g., 2022, 2023).b. Gender: Gender represents a categorical dimension indicating the gender of the students participating in the placement process.c. Specialization: Specialization refers to the academic field or program in which students are enrolled (e.g., engineering, management, computer science).d. ICET Rank or Management Background: This dimension categorizes students based on their performance in the ICET exam or their background in management.e. Company: The company dimension represents the organizations where students receive placement offers.

Measures: Number of Offers: This measure quantifies the total number of placement offers received by students.b. Package Offered: Package offered represents the salary or compensation offered to students by companies during placements.c. Number of Students Placed: This measure counts the total number of students who secured placements.d. Average Package: Average package calculates the mean salary or compensation offered to students across all placements.e. Placement Count: Placement count measures the total number of placement opportunities secured by students.f. Highest Package: This measure identifies the maximum salary or compensation offered to a student during placements.g. Number of Companies: Number of companies represents the count of unique organizations participating in the placement process.h. ICET Rank: ICET rank indicates the rank secured by students in the Integrated Common Entrance Test.

### • Identifying KPIs

Identifying Key Performance Indicators (KPIs) in a the project involves selecting specific metrics that align with project objectives to measure success and performance. This process entails defining relevant criteria, such as placement rates, student satisfaction, or average salary offers, that directly reflect project goals. KPIs provide a quantifiable and actionable framework for evaluating the effectiveness of strategies and interventions. They help stakeholders focus on critical areas, prioritize resources, and make informed decisions based on measurable outcomes. Regular monitoring and adjustment of KPIs ensure ongoing alignment with project objectives and facilitate continuous improvement efforts.

### • ETL using Python

Python provides a rich ecosystem of libraries and tools for performing each phase of the ETL process.

By leveraging Python's versatility, data analysts and engineers can efficiently extract, transform, and load data to support analytics and decision-making processes.

ETL is a fundamental component of the university data analytics project, enabling the transformation of raw data into actionable insights to drive improvements in placement outcomes and strategies.

The ETL process aligns with the project objective of analyzing university placement data to improve outcomes and strategies.

ETL ensures that data is cleansed, integrated, and prepared for analysis, enhancing its usability and reliability.

The transformed and loaded data serves as the foundation for generating insights, identifying trends, and making data-driven decisions.

ETL is an iterative process, allowing for refinements and adjustments based on feedback and evolving project requirements.

1. Extract:

Python Approach:

Python offers several libraries and tools for data extraction:

Pandas: Ideal for reading data from files like CSV, Excel, and databases using read\_csv(), read\_excel(), read\_sql(), etc.

1. Transform:

The transformation phase involves cleaning, filtering, and structuring the extracted data to make it suitable for analysis.

Python Approach:

Python provides powerful libraries for data manipulation and transformation:

Pandas: Offers functions for data cleaning, filtering, grouping, and aggregation.

NumPy: Useful for mathematical operations and array manipulation.

scikit-learn: Provides tools for data preprocessing, feature engineering, and scaling. 3. Load:

The loading phase involves storing the transformed data into a target destination, such as a database, data warehouse, or file.

Python Approach:

Python offers multiple options for loading data into different destinations: Pandas: Writing data to various formats like CSV, Excel, or databases using to\_csv(), to\_excel(), to\_sql(), etc and analyse data and generate insights.

### • Using tools such as Jupyter Notebooks, Power BI and Tableau

In our university data analytics project, Jupyter Notebook serves as a versatile tool for data exploration, analysis, and modeling. It allows us to interactively execute Python code, visualize data, and document our analysis workflows, facilitating collaboration and reproducibility.

Tableau and Power BI are essential for data visualization and dashboarding, enabling us to create interactive and insightful visualizations from our analyzed data. These tools empower stakeholders to explore data trends, identify patterns, and gain actionable insights through intuitive dashboards and reports. Additionally, they support storytelling by effectively communicating findings and recommendations to diverse audiences, enhancing decision-making and driving positive outcomes.

Jupyter Notebook:

Jupyter Notebook serves as an interactive environment for data exploration, analysis, and modeling using Python.

Pandas: For data manipulation, cleaning, and preprocessing.

NumPy: For numerical operations and array manipulations.

scikit-learn: For machine learning tasks such as regression, classification, and clustering.

Matplotlib, Seaborn, Plotly: For data visualization and creating plots, charts, and graphs. Tableau:

Tableau is employed for creating interactive dashboards and visualizations to explore and communicate insights from the data.

Power BI:

Power BI is utilized for building interactive reports and dashboards, similar to Tableau, to visualize and share insights.

In the university data analytics project, Jupyter Notebook is used initially to clean, preprocess, and analyze the placement data using Pandas and NumPy. Machine learning models are built and evaluated using scikit-learn.

Once the data is prepared and insights are derived, the findings are visualized using Tableau or Power BI. Interactive dashboards and visualizations are created to showcase placement trends, student demographics, and company preferences.

By leveraging these tools and packages, the university data analytics project can efficiently analyze, visualize, and communicate insights from the data, enabling stakeholders to make informed decisions and optimize placement strategies.

### • Developing Visualizations as per KPIs

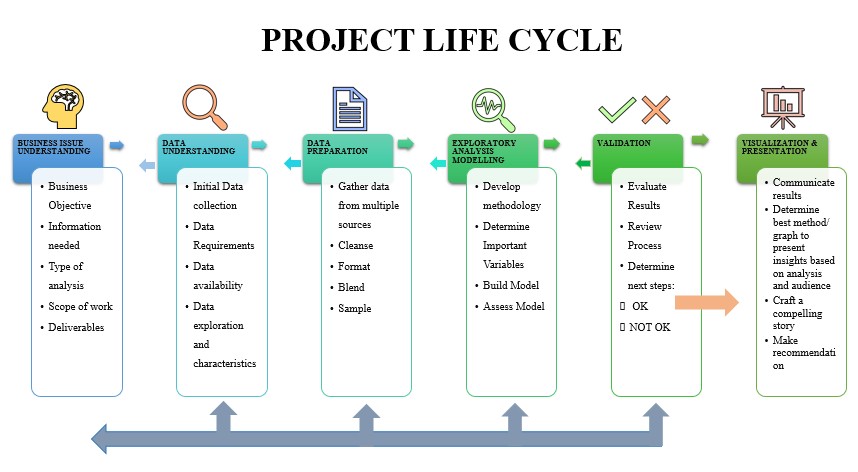
* Explore placement outcomes based on MBA specialization (HR, Finance, Marketing ).
* Identifying which specialization has the maximum number of placements.
* Analyze the relationship between ICET rank and student placements.
* Investigate gender disparities in placement outcomes.
* Determine the number of offers received by students and their distribution across various attributes.
* Analyze placement patterns pertaining to companies.

The KPI’s pertaining to our project are as follows:

* Average package per year
* No of companies
* No of students eligible for placements
* No of students placed
* Maximum package
* No of companies
* Total no of offers
* Placement rate

Based on these KPI’s visualization charts, graphs and dashboards are developed.

### • Understanding of Project Life Cycle



#### 1. Understand the Business Issues

When presented with a data project, you will be given a brief outline of the expectations. From that outline, you should identify the key objectives that the business is trying to uncover. You should examine the overall scope of the work, business objectives, information the stakeholders are seeking, the type of analysis they want you to use, and the deliverables (the outputs of the project) they want.

You need to have these elements clearly defined prior to beginning your data analysis project to provide the best deliverable you can. Additionally, it’s important to ask as many questions as you can at the outset of the project because, often, you may not have another chance before the completion of the project.

#### 2. Understand Your Data Set

There are a variety of tools you can use to organize your data. When presented with a small dataset, you can use Excel, but for heftier jobs, you’ll likely want to use more rigid tools to explore and prepare your data. Muñoz suggests R, Python, Alteryx, Tableau Prep or Tableau Desktop to help prepare your data for it’s cleaning.

Within these programs, you should identify key variables to help categorize the data. When going through the data sets, look for errors in the data. These can be anything from omitted data, data that doesn’t logically make sense, duplicate data, or even spelling errors. These missing variables need to be amended so you can properly clean your data.

#### 3. Prepare the Data

Once you have organized and identified all the variables in your dataset, you can begin cleaning. In this step, you will input missing variables, create new broad categories to help categorize data that doesn’t have a proper place, and remove any duplicates in your data. Imputing average data scores for categories where there are missing values will help the data be processed more efficiently without skewing it.

#### 4. Perform Exploratory Analysis and Modeling

In this step, you will begin building models to test your data and seek out answers to the objectives given. Using different statistical modeling methods, you can determine which is the best for your data. Common models include linear regressions, decision trees, and random forest modeling, among others.

#### 5. Validate Your Data

Once you have crafted your models, you’ll need to assess the data and determine if you have the correct information for your deliverable. Did the models work properly? Does the data need more cleaning? Did you find the outcome the client was looking to answer? If not, you may need to go over the previous steps again.

You should expect a lot of trial and error!

#### 6. Visualize and Present Your Findings

Once you have all your deliverables met, you can begin your data visualization. In many cases, data visualization will be crucial in communicating your findings to the client. Not all clients are data-savvy, and interactive visualization tools like Tableau are tremendously useful in illustrating your conclusions to clients. Being able to tell a story with your data is essential. Telling a story will help explain to the client the value of your findings.

As with any project, you need to identify your objectives clearly. Outlining your work will ensure you get the best deliverables for your clients. While all of these steps are important, if you start the project without all the data you need, you are likely to have to backtrack.

### • Project Documentation

**OBJECTIVE:**

The objective of this project is to develop an integrated data analytics framework for optimizing university operations and enhancing student experiences. This project aims to leverage data-driven insights to improve resource allocation, student retention strategies, academic performance tracking, and institutional decision-making processes. By analyzing various data including relationship between various factors such as student education, family background, academic records, student demographics and university initiatives with student placement outcomes. By identifying key influencers, the project aims to devise targeted strategies to enhance placement rates and better support students in their career aspirations. It aims to provide actionable recommendations that empower university administrators to make informed decisions, foster student success, and streamline operational efficiency. This involves leveraging data-driven insights to improve resource allocation, student retention strategies, academic performance tracking, and institutional decision-making processes. The project addresses critical business needs by enhancing placement success, optimizing curriculum relevance, strengthening industry partnerships, and personalizing career guidance.

**PROBLEM DEFINITION:**

University placement analysis is critical for evaluating the effectiveness of placement strategies and improving student outcomes post-graduation. By analyzing various attributes of students, their education, and the efforts of the placement and training cell, universities can optimize their placement processes and enhance student employability.

This project aims to analyze student placement outcomes based on various attributes such as educational background, specialization, ICET rank, parental occupation, gender, and address type (urban or rural). The goal is to identify patterns and factors influencing student placements to devise targeted strategies for improving placement rates and enhancing student career prospects.

**SUMMARY:**

The analysis aims to improve student placement outcomes at the university by understanding factors influencing success and implementing targeted strategies. This data-driven approach involves leveraging diverse datasets and advanced analytics techniques to identify predictors of placement success and candidate suitability. Key challenges faced include limited visibility into job market trends, insufficient career guidance, mismatched expectations, ineffective skill development, and limited placement opportunities. By addressing these challenges, the university aims to enhance placement success, optimize curriculum relevance, strengthen industry partnerships, and personalize career guidance for students. This involves tailoring placement strategies, aligning curriculum with employer demands, and providing personalized support based on individual profiles. Ultimately, the goal is to improve overall placement rates and support students in achieving their career goals.

**ENTITIES AND SUB-ENTITIES:**

**ENTITIES:**

* Student
* Student Education
* Student placement
* University
* Placement and training cell
* Recruiter/Companies **SUB ENTITIES:**
* Post Graduation
* Degree
* 12 education

**ENTITIES AND THEIR ATTRIBUTES:**

University:

* Students
* Recruiters/Companies • Placement and training cell

Student:

* Hall. Ticket no
* Gender

Student Education:

* ICET rank

MBA:

* Graduation year
* MBA CGPA
* MBA specialization Student placement:
* No of placements
* Placement Status
* LPA

Student Father:

* Father education
* Father occupation
* Father annual income

Placement and training cell:

* Recruitment drives
* Recruiter/Company Recruiter/Company:
* Company name • No of students recruited

Degree:

* Graduation year
* Specialization
* CGPA

12 Education:

* Stream of education
* CGPA
* Year of passing

**DATA MODELLING:**

Data modeling in the context of university placement data analysis project involves structuring and organizing the available data entities and their attributes to facilitate effective analysis and decision-making.

Data modeling entails organizing data entities such as Students,

Recruiters/Companies, and Placement and Training Cell into logical structures.

The relationships between entities, such as the association between Students and Placements, are also modeled.

Data modeling serves as the foundation for organizing and analyzing diverse data elements to derive meaningful insights and enhance placement strategies in the university.

Here, we have depicted the system related Process flow diagram, Use-case diagram, State chart diagram, Entity Relationship diagram and Activity diagrams.

**PROCESS FLOW DIAGRAM:**

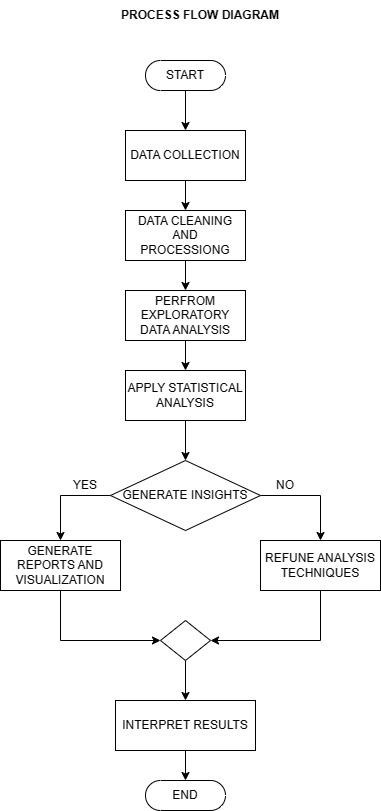


Figure : Process flow diagram of the project

In the process flow diagram of a university placement data analysis project, data collection involves gathering student and placement-related data. Cleaning data includes removing inconsistencies and errors. Data pre-processing prepares data for analysis by normalization and feature engineering. Analyzing data involves statistical techniques to derive insights. Modeling data uses machine learning algorithms for predictive analysis. Visualizing data presents insights through graphs and charts. Refining analysis techniques iteratively enhances data processing methods. Results and insights generation synthesizes findings for actionable outcomes.

**ENTITY RELATIONSHIP DIAGRAM:**

Identifying Entities and Attributes and Attributes for each entity, Identifying relationships between various entities and establishing the relation between different entities. For example, a student is associated with a university, so there is a relationship between Student and University entities. Similarly, a student's placement is associated with a company, indicating a relationship between Student and Recruiters/Companies entities. Using symbols to represent entities, attributes, and relationships. Rectangles represent entities, while ovals represent attributes.

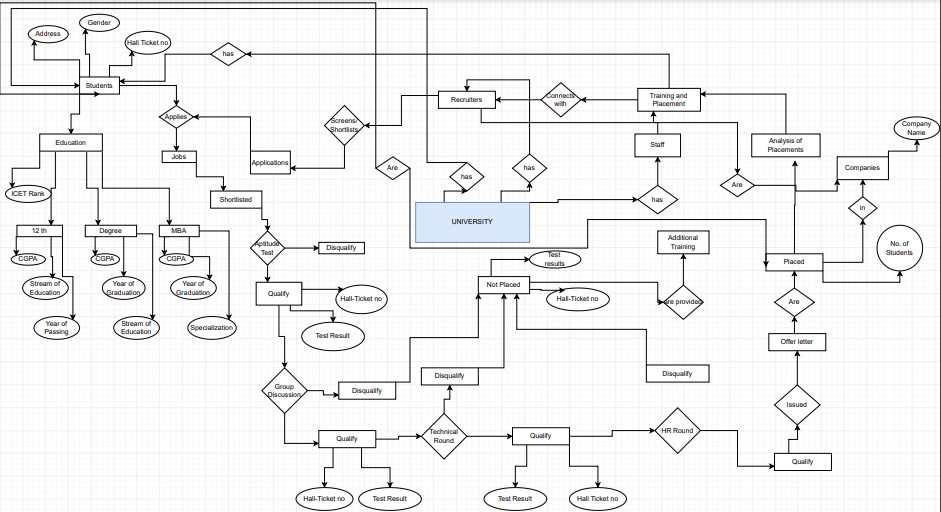


Figure : Entity Relationship diagram of the system

**ACTIVITY DIAGRAM:**

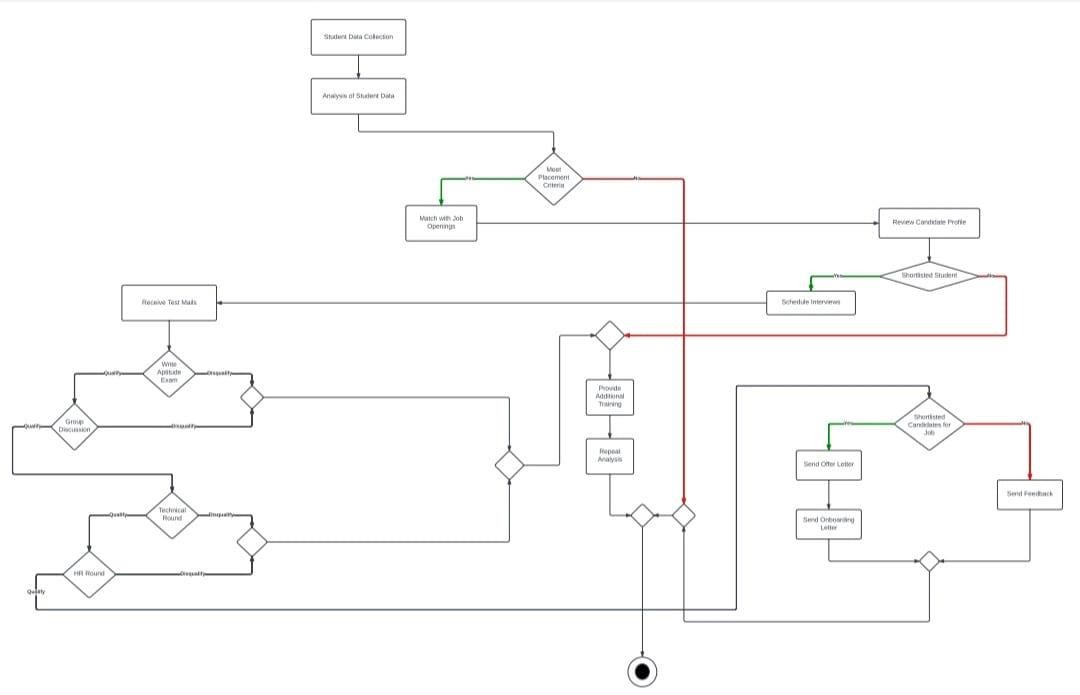


Figure : Activity diagram of the project

In the activity diagram for the university placement data analysis project, the process begins with data collection placement and training cell. The collected data is then preprocessed to clean and prepare it for analysis. Next, data analysis is performed to extract insights such as the number of placements, interview rounds passed, and offer letters received. Finally, visualization techniques are applied to present the analyzed data, aiding in decision-making processes related to onboarding and optimizing recruitment strategies.

**STATE CHART DIAGRAM:**

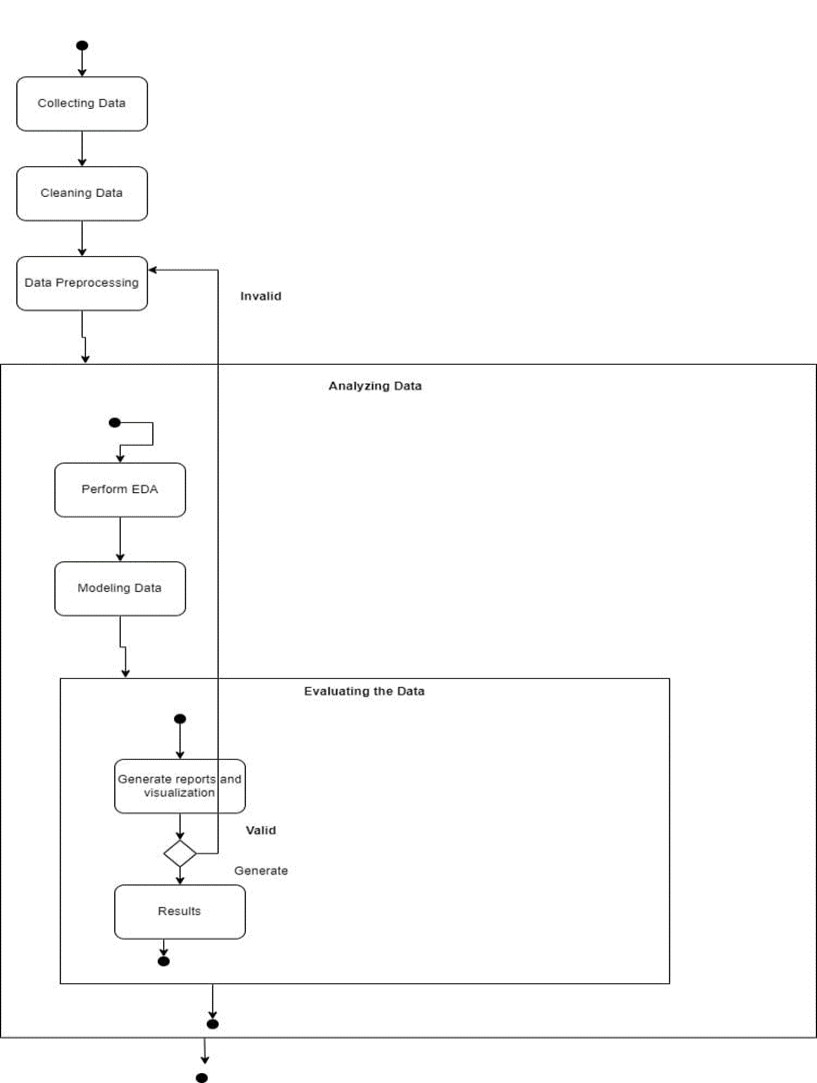


Figure : State chart diagram of the project

The state chart diagram for a university placement data analysis project with the mentioned states can visually represent the flow of the data analysis process. The initial state where data collection processes are initiated. This state involves activities such as gathering student placement records, recruiter information, and other relevant data sources. Once data is collected, it is cleaned where inconsistencies, errors, and missing values are addressed. After cleaning, data undergoes preprocessing to prepare it for analysis.

The pre-processed data is then analyzed to identify patterns, trends, and relationships. Based on the analysis, models are developed to predict placement outcomes or classify students into different categories. Data visualization techniques are applied to present analysis results in a clear and understandable format. The analysis results and insights are generated.

**USE CASE DIAGRAM:**

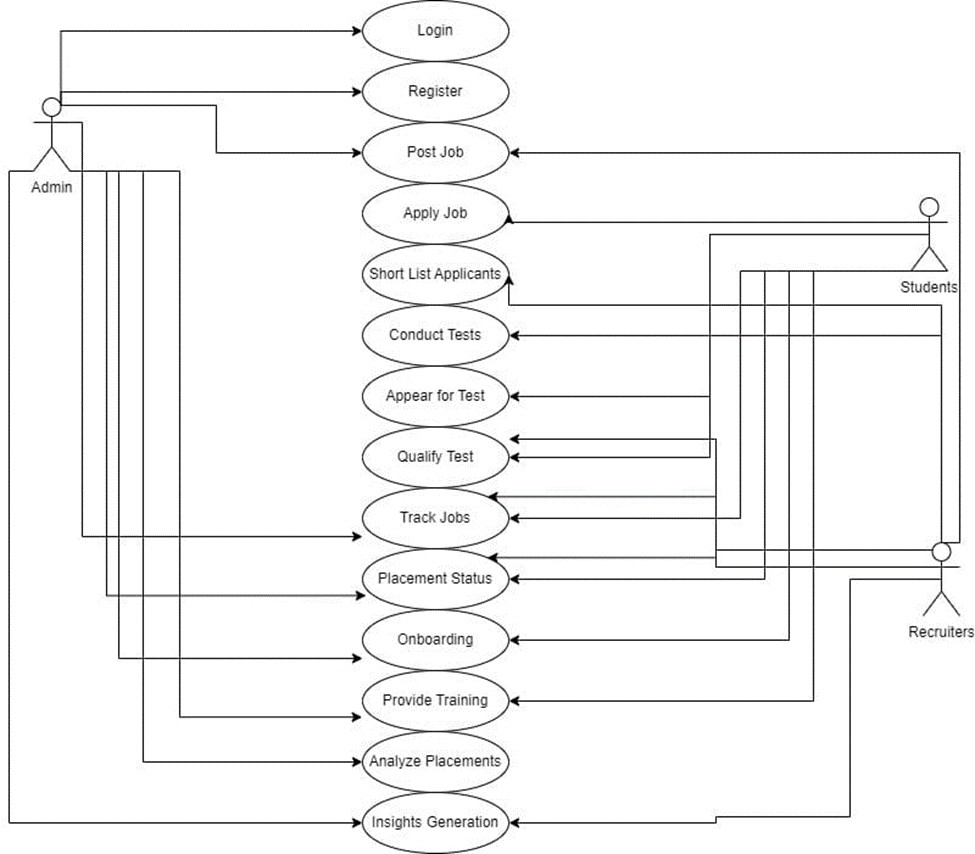


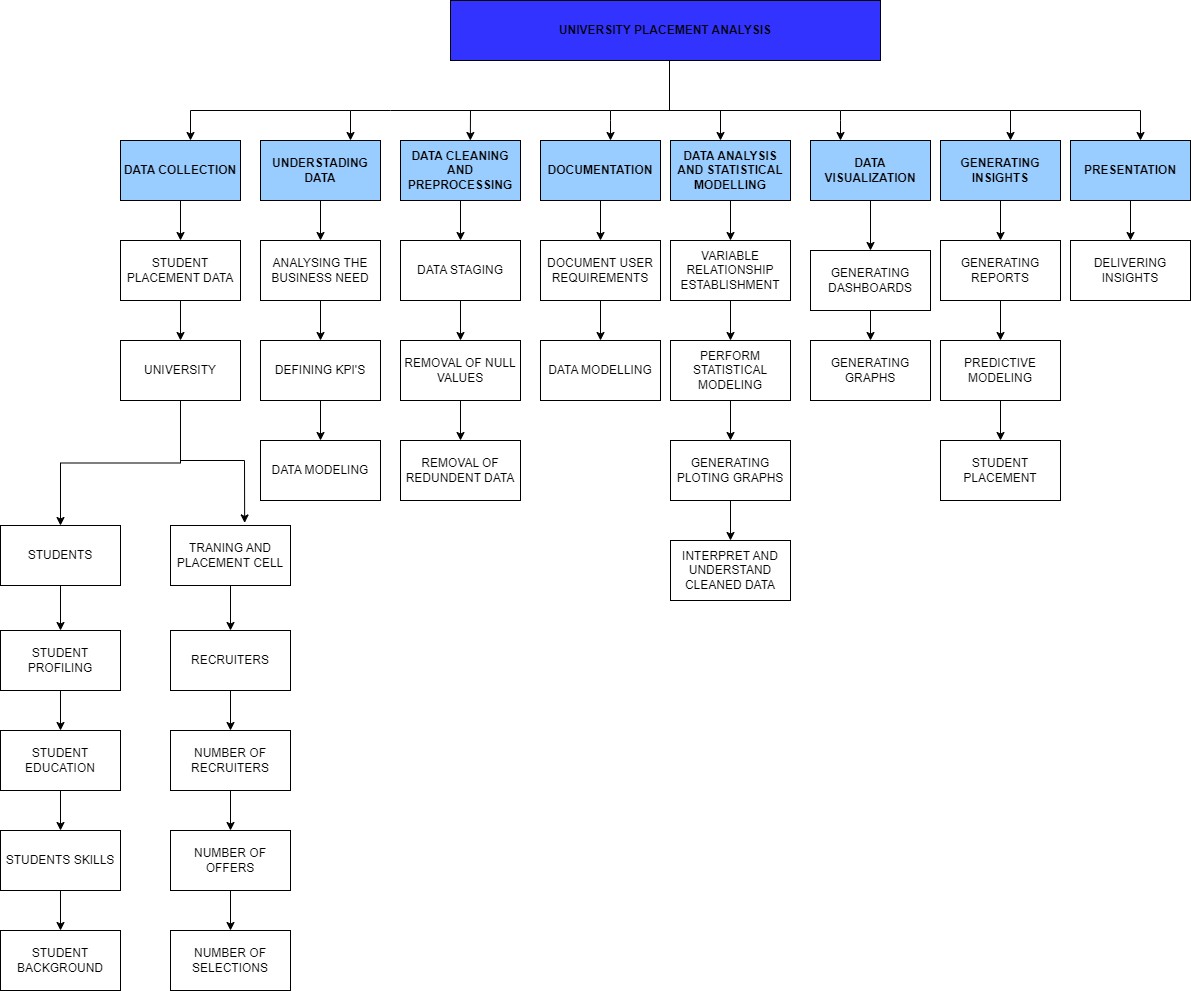
Figure : Use Case diagram of the project

In a university placement data analysis project, the use case diagram illustrates the interactions between users and the system. Admin of university placement and training department can log in, register recruiters/companies, and analyze placements.

Students can log in, view placement details, and accept offer letters. Recruiters/companies can log in, post recruitment drives, and view student profiles. Manage recruitment drives, coordinate with recruiters, and update placement status.

Facilitate interview rounds, track onboarding processes, and generate offer letters.

**MODULES OF PROJECT:**



The university placement analysis project ensures a systematic approach to data analysis, resulting in valuable insights that can inform decision-making and improve placement strategies.

The following are the various phases of the project:

DATA COLLECTION:

Gathering data from various sources such as student records, placement and training cell reports, recruiter/company data, academic performance records, and demographic information. Also involves usage of data collection tools like databases, spreadsheets. Gain a comprehensive view of the data landscape, understand the data sources, and identify potential data gaps or inconsistencies.

UNDERSTANDING THE DATA:

To understand data distributions, patterns, correlations and outliers establish relationships between entities such as students, recruiters/companies, placement and training cell, student education, and family background. Identify key trends, correlations, and potential factors influencing student placements.

DOCUMETATION:

Document data sources, data collection methods, data cleaning and preprocessing steps, and initial findings from the data collection phase to presentation phase.

Use documentation tools like Microsoft Word, Google Docs, or project management platforms with documentation features.

Documenting initial insights, hypotheses, and questions for further analysis.

DATA CLEANING AND PREPROCESSING:

Cleanse data by handling missing values, outliers, data inconsistencies, and formatting issues. Preprocess data for analysis by encoding categorical variables, scaling numerical features, and creating derived features if necessary.

Use data cleaning tools like Python (Pandas, NumPy), R, or data preprocessing libraries such as scikit-learn. Ensure data consistency and accuracy for reliable analysis and modeling.

Identify data quality issues and improve data readiness ensuring that cleaned and preprocessed dataset ready for modeling and analysis.

DATA ANALYSIS AND STATISTICAL MODELLING:

Statistical analysis, machine learning algorithms, and other analytical techniques are applied to the prepared dataset to extract insights and identify patterns.

Analysis techniques include exploratory data analysis, hypothesis testing, regression analysis, classification, clustering, and more, depending on the project objectives.

Statistical models are developed to predict placement outcomes, identify significant factors influencing placements, or forecast future trends.

Modeling techniques such as linear regression, logistic regression, decision trees, random forests, and others are employed to analyze the relationship between variables and make predictions.

DATA VISUALIZATION:

Create visualizations such as bar charts, histograms, scatter plots, heatmaps, and dashboards to represent data relationships, trends, and patterns.

Use statistical analysis tools like Python (Pandas, NumPy), R, or data visualization tools like Tableau, Power BI, or matplotlib/seaborn for visualization. Visualize relationships between variables, distributions, and trends over time or across categories.

GENERATING INSIGHTS:

Apply statistical analysis, machine learning algorithms, and data mining techniques to generate actionable insights report with recommendations for decisionmaking.

Use Python (scikit-learn, statsmodels), R, or specialized analytics platforms for advanced analysis. Uncover relationships, correlations, predictive models, and factors influencing student placements.

Derive insights such as key predictors of placement success, demographic trends, and performance metrics.

PRESENTATION:

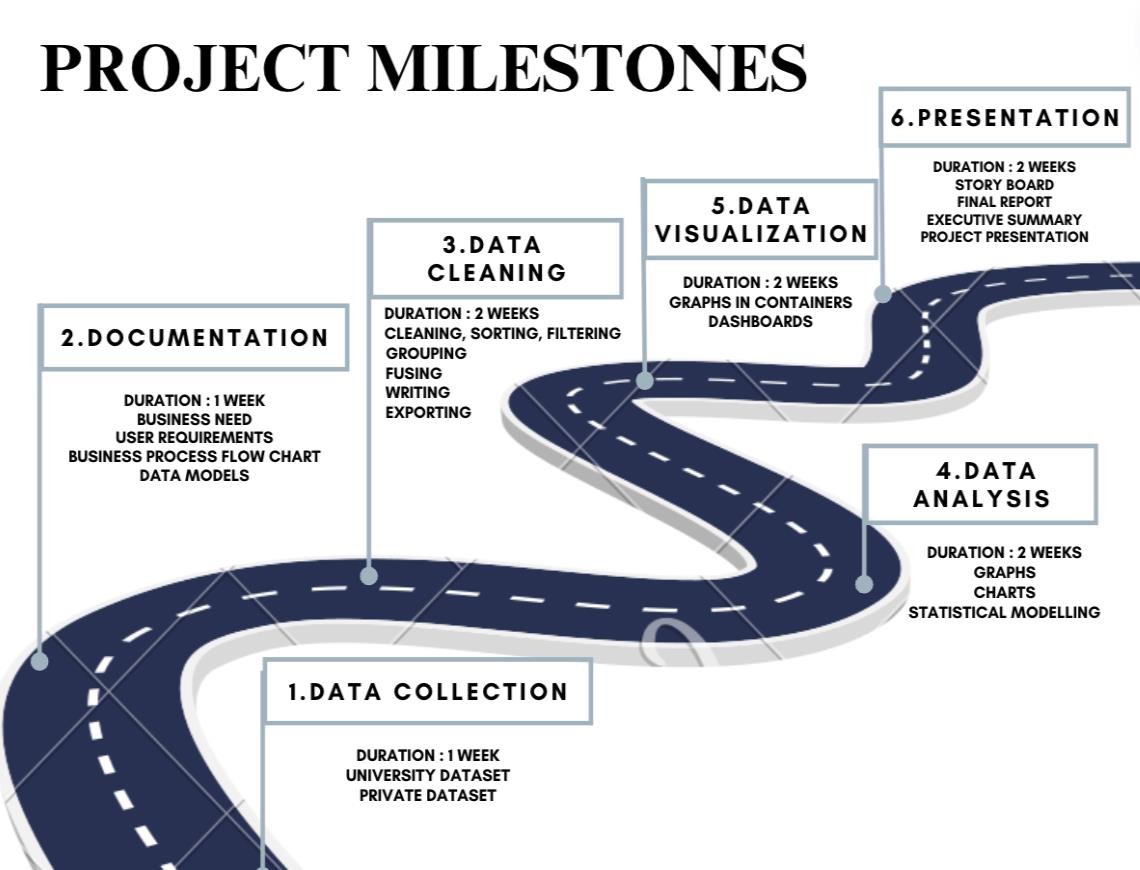
Create a comprehensive presentation summarizing the project phases, key findings, insights, recommendations, and future steps.

Usage of presentation tools like Microsoft PowerPoint, Google Slides, or visualization platforms for interactive presentations. Presenting relationships between data analysis results, strategic recommendations, and project objectives.

Communicate insights clearly, highlighting the key findings and trends and impact on university operations and student outcomes effectively through visualizations.

# Project Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| SNO | Milestone | Description | Duration |
| 1. | DATA  COLLECTION | Data Collection:  Gather data from university and private datasets. | 1 week |
| 2. | DOCUMENTATION | Documentation:  Identify business needs, document user requirements, create process flow charts, and develop data models. | 1 week |
| 3. | DATA CLEANING | Data Cleaning:  Clean, sort, filter, group, fuse, and write data to prepare it for analysis. | 2 weeks |
| 4. | DATA ANALYSIS | Data Analysis:  Create graphs, charts, and perform statistical modeling to interpret and understand the cleaned data. | 2 weeks |
| 5. | DATA  VISUALIZATION | Data Visualization:  Develop visual representations (graphs, dashboards) to make insights more understandable. | 2 weeks |
| 6. | PRESENTATION | Presentation:  Prepare a storyboard, finalize the project report (including an executive summary), and deliver the project presentation. | 2 weeks |



1. Data Collection:

Duration: One week

Tasks: Gather data from university and private datasets.

Tools:

Excel: For basic data collection and initial organization.

1. Documentation:

Duration: One week

Tasks: Identify business needs, document user requirements, create process flow charts, and develop data models.

Tools:

Microsoft Word for documenting user requirements and project

details.

Lucidchart, Draw.io,Star uml for creating process flowcharts.

1. Data Cleaning:

Duration: Two weeks

Tasks: Clean, sort, filter, group, fuse, and write data to prepare it for analysis.

Tools:

Excel: for cleaning and data staging

Python libraries like Pandas: For data cleaning and manipulation.

1. Data Analysis:

Duration: Two weeks

Tasks: Create graphs, charts, and perform statistical modeling to interpret and understand the cleaned data.

Tools:

Jupyter Notebook: For interactive data analysis and visualization using Python.

Python libraries like Matplotlib, Seaborn, Sklearn and Plotly for data visualization.

Perform statistical modelling on data using Statistical methods : LINEAR

REGRESSION, ANOVA, CLUSTER ANALYSIS.

1. Data Visualization:

Duration: Two weeks

Tasks: Develop visual representations graphs, dashboards in Tableau and Power BI to make insights more understandable.

Tools:

Tableau: For creating interactive dashboards and visualizations.

Power BI and Tableau: For building interactive reports and dashboards.

Python libraries like Matplotlib, Seaborn, and Plotly for creating static and interactive visualizations.

1. Presentation:

Duration: Two weeks

Tasks: Prepare a storyboard, finalize the project report (including an executive summary), and deliver the project presentation.

Tools:

Microsoft PowerPoint for creating the presentation slides.

# Requirements Specifications

* **User Requirements**
* No. of Companies: This metric indicates the total number of companies participating in the placement process. It reflects the diversity of employers engaging with the university.
* Overall Placed: This figure represents the total count of students successfully placed in jobs or internships. It provides an overview of the placement success rate.
* Average Package per Year: The average annual salary package secured by placed students. It reflects the compensation level offered by employers.
* Specialization by No. of Offers: This point likely refers to the distribution of placements across different academic specializations (e.g., engineering, business, etc.). It highlights which fields have more job opportunities.
* Average Package Specialization: The average salary package specific to each specialization. It helps assess the earning potential in different fields.
* Highest Package in Specialization: The maximum salary package offered within a specific specialization. It showcases the top-tier opportunities available.
* Male & Female Placed: This point likely indicates the gender distribution among placed students. It’s essential for assessing gender equity in placements.
* No. of Offers: The total number of job offers extended to students. It reflects the demand for talent from employers.
* No. of Companies: This metric indicates the total number of companies participating in the placement process. It reflects the diversity of employers engaging with the university.
* Max Package: This likely represents the highest salary package offered or received by students. It showcases the top-tier opportunities available.
* Minimum Package: Indicates the smallest or lowest salary package available. It helps assess the baseline compensation level.
* Repeated Companies by Package: This point may refer to companies that appear multiple times in the placement data due to varying salary packages. It highlights variations within the same company.
* Top 5 Companies by Package: Displays the five leading companies based on their salary packages. It identifies the most attractive employers for students.
* Top 5 Companies by Offers: Shows the top five companies in terms of the number of job offers extended. It reflects demand and popularity among students.
* **Functional Specifications**

User/client requirements for a university data analysis project outline the specific needs, expectations, and functionalities desired by the end-users or clients who will interact with the system or software. These requirements are crucial for ensuring that the final product meets the intended use case and provides value to stakeholders. Here's an explanation of user/client requirements for the university data analysis project:

Data Accessibility and Integration:

Users require access to comprehensive and up-to-date datasets from various sources, including university databases, placement records, surveys, and external datasets. They expect the system to seamlessly integrate and consolidate heterogeneous data formats and structures, ensuring accessibility and usability.

Intuitive User Interface:

Users expect an intuitive and user-friendly interface that allows for easy navigation and interaction with the system.

They require clear and organized layouts, menus, and controls to facilitate efficient data exploration, analysis, and visualization.

Customization and Flexibility:

Users may have diverse needs and preferences regarding data analysis and visualization.

They require the system to offer customization options such as customizable dashboards, visualizations, and report templates to tailor the user experience to their specific requirements.

Analytical Capabilities:

Users expect the system to provide robust analytical capabilities for exploring, analyzing, and interpreting data.

They require access to a wide range of statistical tools, machine learning algorithms, and visualization techniques to derive meaningful insights and make informed decisions.

Collaboration and Sharing:

Users often collaborate with colleagues, peers, and stakeholders on data analysis projects.

They require features for sharing data, analyses, and insights, as well as collaborative tools for commenting, version control, and real-time collaboration.

Performance and Reliability:

Users expect the system to perform reliably and efficiently, even when handling large datasets and complex analytical tasks.

They require responsive performance, minimal latency, and high availability to support their workflows and decision-making processes.

Security and Privacy:

Users are concerned about the security and privacy of sensitive data and personal information.

They expect the system to implement robust security measures such as encryption, access controls, and compliance with privacy regulations to safeguard data integrity and confidentiality. Training and Support:

Users may require training and support to effectively use the system and leverage its full capabilities.

• **System Specifications**

System specifications refer to the detailed technical requirements and configurations of a computer system necessary to run specific software applications or perform certain tasks effectively. These specifications encompass hardware and software components and are essential for ensuring optimal performance and compatibility. Here's an explanation of system specifications:

Hardware Specifications:

Processor (CPU): Specifies the type, model, and speed of the central processing unit (CPU). Higher processing power enables faster execution of computational tasks. Memory (RAM): Indicates the amount of random-access memory (RAM) available for temporary data storage and processing. Sufficient RAM is essential for multitasking and handling large datasets efficiently.

Storage (Hard Drive): Specifies the storage capacity and type of storage device (e.g., HDD, SSD). Sufficient storage space is necessary for storing datasets, software applications, and operating system files.

Graphics Processing Unit (GPU): Specifies the type and model of the graphics card, important for tasks involving graphical rendering and computation (e.g., data visualization, image processing).

Software Specifications:

Operating System (OS): Specifies the type and version of the operating system required by the software application. Common OS options include Windows, macOS, and Linux distributions.

Application Software: Specifies the specific software applications or tools needed to perform tasks. Data analysis projects may requires software such as Pytho, Tableau, or Power BI.

Dependencies and Libraries: Specifies any additional software dependencies or libraries required by the application for proper functioning. For example, Python-based applications may require libraries like Pandas, NumPy, or scikit-learn for data analysis tasks.

Tools for data modelling : Lucid chart, Star uml, Draw.io

# Business Process Chart

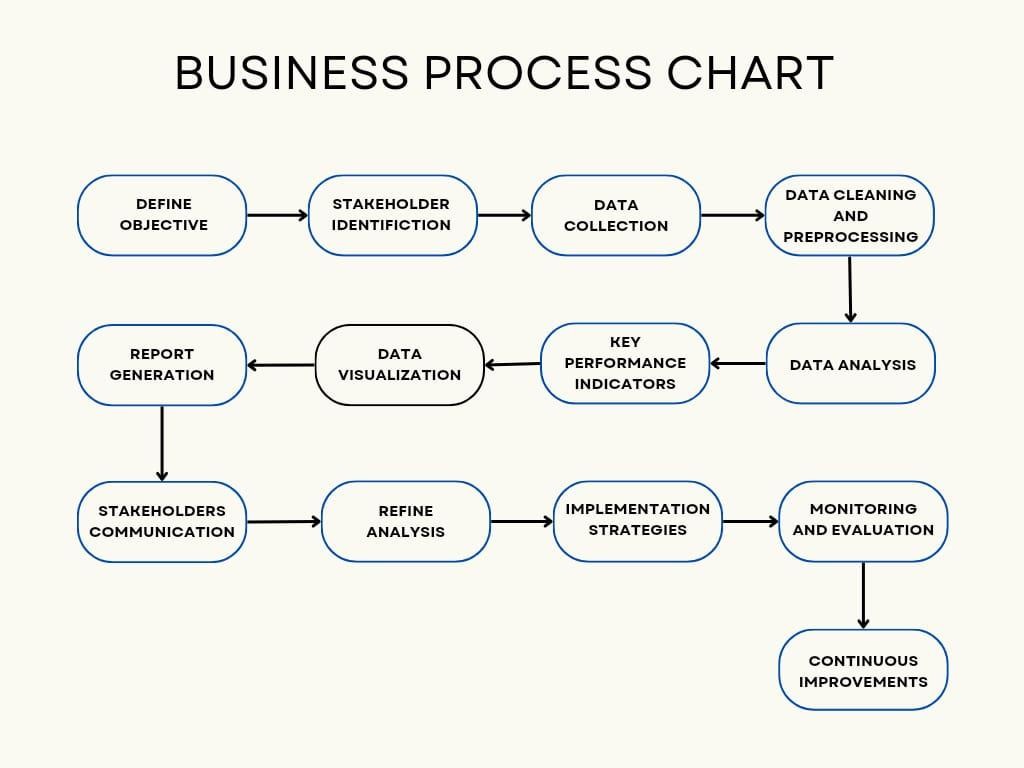


Figure : Business process chart

1. Define Objective:

Objective: Clearly define the goals and objectives of the project, such as improving placement rates, understanding factors influencing placement outcomes, or optimizing placement strategies.

1. Stakeholder Identification:

Identify all stakeholders involved in the placement process, including students, university placement officers, recruiters/companies, alumni, and university administrators.

1. Data Collection:

Collect relevant data from various sources such as university databases, placement records, surveys, and external datasets. Include data on student demographics, academic performance, company profiles, placement outcomes, and more.

1. Data Cleaning and Preprocessing:

Clean and preprocess the collected data to ensure accuracy, consistency, and completeness. Handle missing values, remove duplicates, standardize formats, and encode categorical variables as necessary.

1. Data Analysis:

Perform exploratory data analysis (EDA) to uncover patterns, trends, and correlations in the placement data. Utilize statistical analysis techniques, hypothesis testing, and machine learning algorithms to derive insights and identify factors influencing placement outcomes.

1. KPIs (Key Performance Indicators):

Define key performance indicators (KPIs) to measure the success and effectiveness of placement strategies. KPIs may include placement rates, average package offered, placement success rates by specialization, company-wise placement statistics, and more.

1. Data Visualization:

Create visualizations such as charts, graphs, and dashboards to communicate key findings, trends, and insights effectively to stakeholders. Visualizations facilitate easier interpretation of complex data and support data-driven decision-making.

1. Report Generation:

Generate comprehensive reports summarizing the analysis findings, insights, and recommendations. Reports should be tailored to the needs of different stakeholders and include actionable insights to inform decision-making.

1. Stakeholder Communication:

Communicate analysis findings, insights, and recommendations to stakeholders through presentations, meetings, and workshops. Foster collaboration and engagement among stakeholders to ensure buy-in and alignment with project objectives.

1. Refine Analysis:

Continuously refine and iterate on the analysis based on feedback from stakeholders, validation of insights, and emerging trends. Adjust analysis techniques, data models, and strategies as necessary to improve the accuracy and relevance of insights generated.

1. Implementation Strategies:

Develop and implement strategies based on analysis findings and recommendations to enhance placement outcomes. Strategies may include targeted skill development programs, industry collaborations, personalized career counseling, and more.

1. Monitoring and Evaluation:

Monitor the implementation of strategies and evaluate their impact on placement outcomes. Continuously track KPIs, measure progress against objectives, and identify areas for improvement.

1. Continuous Improvement:

Foster a culture of continuous improvement by learning from successes and failures, adapting to changing circumstances, and refining strategies based on ongoing evaluation and feedback.

# Activities

Define project objectives.

Identify key stakeholders and their requirements.

Establish timelines, milestones, and resource allocations.

Data Collection:

Gather data from university databases, placement records and sources.

Ensure data integrity, completeness, and quality.

Data Cleaning and Preprocessing:

Cleanse data by handling missing values, duplicates, and errors.

Standardize formats, and resolve inconsistencies.

Exploratory Data Analysis (EDA):

Explore data distributions, patterns, and relationships.

Identify outliers, anomalies, and potential insights.

Feature Engineering:

Create new features or variables derived from existing data.

Perform transformations, scaling, or encoding as needed.

Data Modeling and Analysis:

Apply statistical methods, ML algorithms, or predictive models to analyze data.

Evaluate model performance and interpret results.

Data Visualization:

Create visualizations such as charts, graphs, and dashboards to communicate insights.

Use tools like Matplotlib, Seaborn, Tableau, or Power BI for visualization.

Insights Generation:

Derive actionable insights and recommendations from data analysis.

Identify trends, correlations, and factors influencing outcomes.

Report and Presentation:

Prepare comprehensive reports summarizing findings.

Create presentations to communicate results to stakeholders.

Implementation and Deployment:

Implement recommendations or interventions based on insights.

Deploy analytical solutions or tools for ongoing monitoring and evaluation.

Monitoring and Evaluation:

Monitor key metrics and KPIs to track performance over time.

Evaluate the impact of implemented strategies and interventions.

Documentation and Knowledge Sharing:

Document processes, methodologies, and findings for future reference.

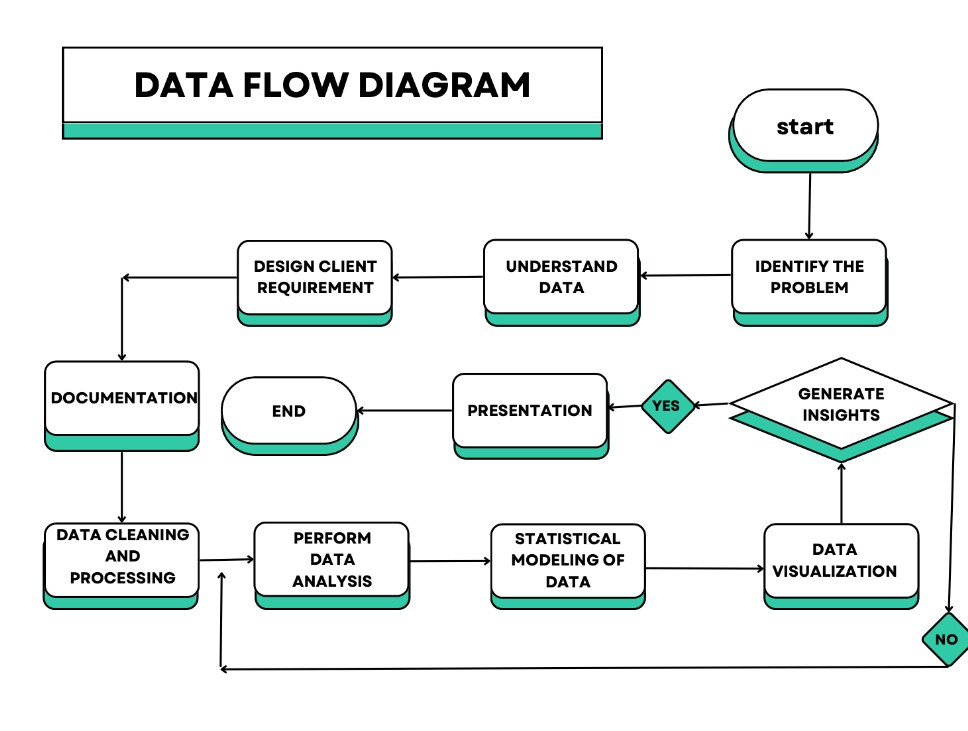
Share knowledge and insights with team members and stakeholders.

Continuous Improvement:

Iterate analysis methodologies, models, and strategies based on feedback & new data.

Incorporate lessons learned and best practices into future projects.

## Data Flow



This data flow diagram outlines the sequential steps involved in the university data analytics project, from data collection to presentation of insights, emphasizing the iterative nature of the analysis process and the importance of refining techniques based on feedback and validation of insights.

Data Collection:

Data sources such as university databases, placement records, surveys, and external datasets feed into the system.

The collected data includes placement records, student information, company details, and other data relevant data for analysis.

Identify the Problem:

This step involves identifying key challenges or objectives the project aims to address, such as improving placement rates, understanding factors influencing placement outcomes, or optimizing placement strategies.

The analysis seeks to identify factors contributing to placement success or failure and propose solutions to enhance placement outcomes.

Understand the Data:

Data exploration techniques are employed to gain a comprehensive understanding of the dataset, including its structure, quality, and potential biases.

Understanding the dataset involves examining placement records, student demographics, company profiles, and other relevant information to inform subsequent analysis.

Design Data Requirements:

Based on the identified problem and understanding of the data, specific data requirements and objectives for the analysis are defined.

Data requirements may include selecting relevant variables, defining performance metrics, and outlining the scope of analysis to achieve the project objectives.

Documentation:

Detailed documentation of the project objectives, data sources, methodology, and analysis plan is created.

Documentation ensures transparency, reproducibility, and clarity throughout the project lifecycle, facilitating collaboration and communication among team members.

Data Cleaning and Processing:

Data preprocessing techniques such as handling missing values, removing duplicates, standardizing formats, and encoding categorical variables are applied to prepare the data for analysis.

Clean and processed data is essential for accurate analysis and modeling, ensuring the reliability and validity of insights generated.

Performing Analysis:

Statistical analysis, machine learning algorithms, and other analytical techniques are applied to the prepared dataset to extract insights and identify patterns.

Analysis techniques include exploratory data analysis, hypothesis testing, regression analysis, classification, clustering, and more, depending on the project objectives.

Statistical Modeling of Data:

Statistical models are developed to predict placement outcomes, identify significant factors influencing placements, or forecast future trends.

Modeling techniques such as linear regression, logistic regression, decision trees, random forests, and others are employed to analyze the relationship between variables and make predictions.

Data Visualization:

Visualization tools and techniques are used to represent the analyzed data in graphical or interactive formats, making complex patterns and trends easier to understand.

Visualizations include charts, graphs, heatmaps, and interactive dashboards to communicate key findings, trends, and insights effectively to stakeholders.

Generating Insights:

Insights are derived from the analyzed data, highlighting actionable recommendations, trends, correlations, and key findings.

Insights include identifying factors influencing placement outcomes, trends in company preferences, opportunities for improvement, and strategies to enhance placement success.

Presentation / Refine Analysis Techniques and Perform Analysis Again:

If insights are valid, they are presented to stakeholders through reports, presentations, or interactive dashboards. If insights are not valid or require further refinement, analysis techniques are adjusted, and the analysis is performed again iteratively.

Stakeholders are informed of the findings and recommendations based on the analysis. Refinement of analysis techniques ensures the accuracy and relevance of insights generated, leading to informed decision-making and continuous improvement in placement strategies.

## Data Models

Data modeling is essential for structuring, analyzing, and interpreting placement data in the university data analytics project. By building data models, we can uncover insights, predict future trends, optimize placement strategies, and support evidence-based decision-making to enhance student outcomes and placement success.

In the context of the university placement data analytics project, data modeling plays a crucial role in structuring and analyzing the placement data to derive actionable insights. Here's an explanation of the concept and importance of data modeling in this project:

Concept of Data Modeling deals with:

Structuring Data: Data modeling involves organizing the placement data into a structured format that facilitates analysis. This includes defining the relationships between different variables or attributes, identifying key entities, and specifying the rules for storing and manipulating data.

Abstraction: Data modeling abstracts complex real-world phenomena into a simplified representation that can be easily understood and analyzed. It allows us to focus on relevant aspects of the data while hiding unnecessary details, making it easier to extract meaningful insights.

Predictive Capabilities: Data modeling enables the development of predictive models that forecast future placement outcomes based on historical data. By analyzing patterns and relationships in the data, we can identify factors influencing placement success and predict future trends, helping to optimize placement strategies.

Optimizing Performance: Effective data modeling can improve the performance of analytical processes by optimizing data storage, retrieval, and manipulation. By designing efficient data structures and algorithms, we can reduce computational overhead and enhance the speed and scalability of analysis.

Facilitating Decision-Making: Data modeling provides a structured framework for organizing and interpreting placement data, enabling stakeholders to make informed decisions based on reliable insights. It helps placement officers, university administrators, employers, and students understand placement trends, identify opportunities for improvement, and devise effective strategies.

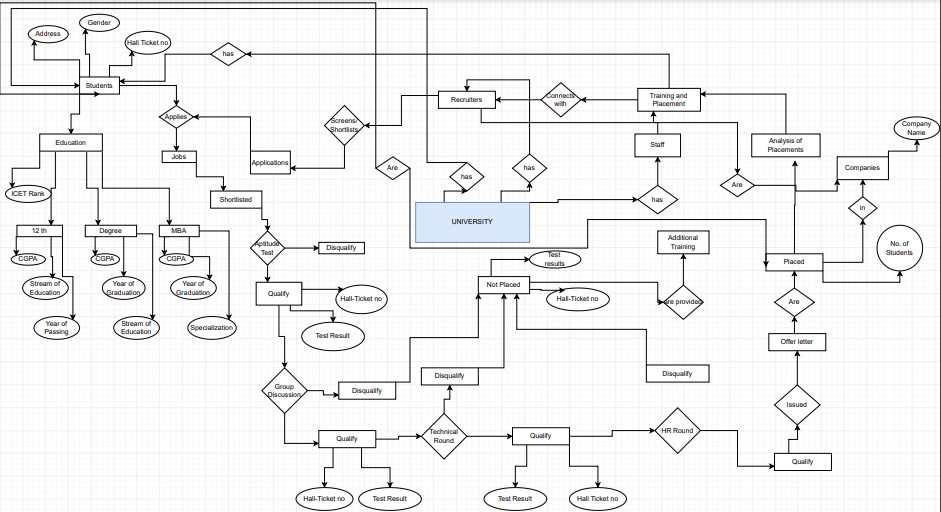
Data models help uncover complex relationships between different variables in the placement dataset. By structuring the data into meaningful entities and defining their relationships, we can analyze how factors such as student demographics, academic performance, company preferences, and placement outcomes are interrelated.

Data models provide a foundation for creating visualizations that effectively communicate insights to stakeholders. By structuring the data into logical entities and relationships, we can generate visualizations such as charts, graphs, and dashboards that highlight key trends, patterns, and correlations in the placement data.

Data modeling is an iterative process that allows us to refine and improve our understanding of the placement data over time. By continuously updating and refining data models based on new insights and feedback, we can adapt to changing placement dynamics, address emerging challenges, and optimize placement strategies for ongoing success.

Data models serve as a foundation for evidence-based decision-making in the placement process. By providing a structured framework for analyzing placement data, data models enable stakeholders to make informed decisions, evaluate the effectiveness of placement initiatives, and drive continuous improvement in placement outcomes.

ER DIAGRAM (ENTITY RELATIONSHIP DIAGRAM)



In the university placement data analysis project, the Entity-Relationship (ER) diagram visually represents the relationships between these entities and sub-entities, facilitating the understanding and analysis of the placement data in the university context.

Entities: Student:

* This entity represents individual students participating in the placement process.
* Attributes include Hall Ticket Number, Gender, and Address.

Student Education:

* This entity captures the educational background of students.
* Sub-entities include ICET Rank, MBA, Post Graduation, Degree, and 12

Education.

Student Placement:

* This entity tracks placement-related information for students. - Attributes include Number of Placements, Placement Status, and LPA

Student Father:

* This entity stores information about students' fathers.
* Attributes include Father Education, Father Occupation, and Father Annual

Income.

University:

* This entity represents the university conducting placement activities. - Attributes may include details about the number of students, recruiters/companies, and the placement and training cell.

Placement and Training Cell:

* This entity manages placement-related activities within the university.
* Attributes include Recruitment Drives and links to Recruiters/Companies.

Recruiter/Companies:

* This entity represents organizations offering placement opportunities to students.
* Attributes include Company Name and Number of Students Recruited.

Sub-Entities:

MBA:

* This sub-entity represents the Master of Business Administration (MBA) degree details, including Graduation Year, CGPA, and Specialization.

Post Graduation:

* This sub-entity captures details about students' post-graduation education, if applicable.

Degree:

* This sub-entity represents undergraduate degree details such as Graduation

Year, Specialization, and CGPA.

12 Education:

* This sub-entity stores information about students' education at the 12th grade level, including Stream of Education, CGPA, and Year of Passing.

Entities and Their Attributes:

University:

* Attributes: Students, Recruiters/Companies, Placement and Training Cell.

Student:

* Attributes: Hall Ticket Number, Gender, Address.

Student Education:

* Attributes: ICET Rank.

MBA:

* Attributes: Graduation Year, MBA CGPA, MBA Specialization.

Student Placement:

* Attributes: Number of Placements, Placement Status, LPA.

Recruiter/Company:

* Attributes: Company Name, Number of Students Recruited.

Degree:

* Attributes: Graduation Year, Specialization, CGPA.

12 Education:

- Attributes: Stream of Education, CGPA, Year of Passing.

# Datasets

The data set we worked on is Anurag University’s MBA department placement data.

We received 2 excel workbooks from our client which are 2022 and 2023 year passout student data.

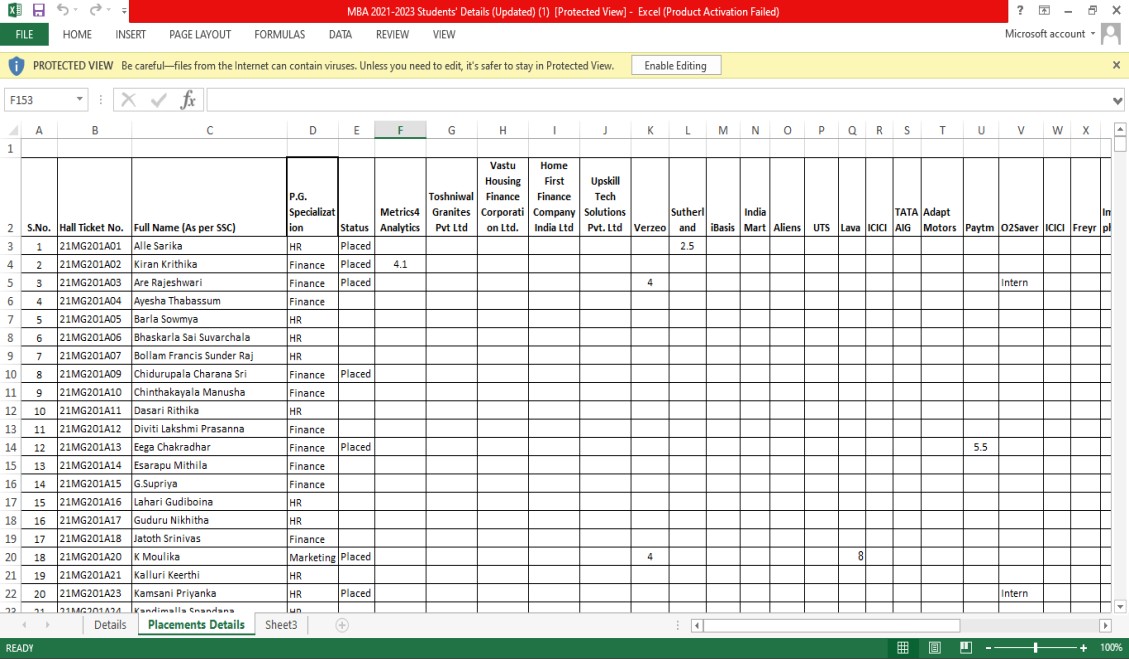
The dataset provides comprehensive information about students, including personal details, educational background, and placement status. It includes details such as names, contact information, educational qualifications, specialization, placement status, and offers received from various companies.

The dataset enables analysis of factors influencing placement outcomes, identification of trends, and formulation of strategies to improve placement rates.

Understanding this dataset is crucial for conducting thorough analysis and deriving actionable insights to enhance the university's placement process.

The dataset comprises two main sheets: "Student Details" and "Student Placement." These sheets contain information about students' personal details, educational background, and placement status. Let's break down the dataset for each year (2022 and 2023) and provide an explanation of the columns in detail:

Dataset for 2023 Passout Student Data:



Student Details Sheet: S.No.: Serial number.

Hall Ticket No.: Unique identification number assigned to each student. Full Name (As per SSC): Student's full name as per their Secondary School

Certificate (SSC).

Father's Name (As per SSC): Father's full name as per SSC.

Father's Education: Educational qualification of the student's father.

Father's Occupation: Occupation of the student's father.

Mother's Name: Mother's full name.

Mother's Education: Educational qualification of the student's mother.

Mother's Occupation: Occupation of the student's mother.

Date of Birth: Student's date of birth.

Gender: Gender of the student.

Caste: Caste of the student.

Caste Sub Category: Sub-category of the student's caste.

Annual Family Income (In LPA): Annual family income in Lakhs per Annum.

Mobile No.: Student's mobile number.

Alternate no.: Alternative contact number.

Email Address (College Mail Id): Student's email address using the college domain.

Alternative Mail ID: Alternative email address.

Aadhar: Aadhar card number.

Date of Joining (DD-MM-YYYY): Date when the student joined the university.

ICET Rank: Rank secured by the student in the Integrated Common Entrance

Test (ICET).

Address: Student's address.

Languages Known: Languages known by the student.

Mother Tongue: Student's mother tongue.

Foreign Language(s) Known: Foreign languages known by the student.

10th %: Percentage obtained in 10th grade.

10th Board: Board of examination for 10th grade.

10th Passing Year: Year of passing 10th grade.

12th %: Percentage obtained in 12th grade.

12th Board: Board of examination for 12th grade.

12th Branch Name: Stream/branch of 12th grade.

12th Passing Year: Year of passing 12th grade.

Graduation %: Percentage obtained in graduation.

Graduation Branch: Branch of graduation.

Graduation University: University from which graduation was completed.

Graduation Passing Year: Year of passing graduation.

Post Graduation: Indicates if the student pursued post-graduation. P.G. Specialization: Specialization pursued in post-graduation.

Status: Current status of the student.

P.G. Backlogs: Backlogs in post-graduation, if any.

Current Residence: Student's current place of residence.

Interested Job Role: Job role the student is interested in.

Expecting Salary: Expected salary by the student.

Preferred Location to work: Preferred work location by the student.

Willing to relocate to work location: Indicates if the student is willing to relocate for work.

Need Training?: Indicates if the student needs training.

Willing to take certification courses: Indicates if the student is willing to take certification courses.

Upload CV (with Photo) (in PDF format only): CV uploaded by the student.

If Need Training, which is area you need training?: Area in which the student needs training.

Student Placement Sheet: S.No.: Serial number.

Hall Ticket No.: Unique identification number assigned to each student. Full Name (As per SSC): Student's full name as per their Secondary School

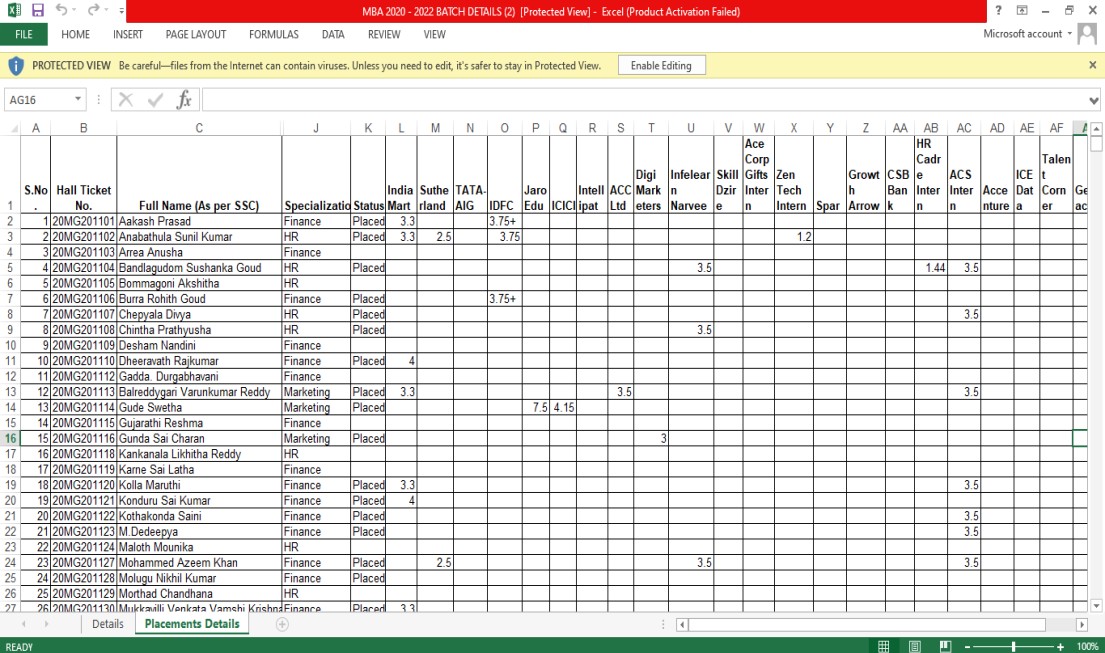
Certificate (SSC).

P.G. Specialization: Specialization pursued in post-graduation.

Status: Current status of the student.

Metrics4 Analytics to Total Offers: Indicates the number of offers received from various companies.

Dataset for 2022 Passout Student Placement Data:



Student Details Sheet:

SL.NO.: Serial number.

Hall Ticket No.: Unique identification number assigned to each student.

Name of the Candidate (With Surname): Student's full name.

Phone Number: Student's phone number.

I-CET RANK: Rank secured by the student in the Integrated Common Entrance

Test (ICET).

Mentor: Mentor assigned to the student.

Date Of Joining: Date when the student joined the university.

Father Name: Father's full name.

Father Ph.No: Father's phone number.

Occupation: Occupation of the student's father.

Mother Name: Mother's full name.

Mother Ph.No: Mother's phone number.

Degree % & Stream: Percentage obtained and stream of the student's degree.

Degree Y.O.P: Year of passing degree.

Degree College: College from which the degree was obtained.

Inter % & Stream: Percentage obtained and stream of intermediate education.

Inter Year of Pass: Year of passing intermediate education.

Intermediate College: College from which intermediate education was obtained.

SSC %: Percentage obtained in Secondary School Certificate (SSC).

SSC Y.O.P: Year of passing SSC.

Schooling: Schooling details.

Aadhar Number: Aadhar card number.

Email-id: Student's email address.

College Mail ID: Student's email address using the college domain.

Date .Of .Birth: Student's date of birth.

Seat Category: Category of seat allocated.

Special Interests: Special interests of the student.

Student Placement Sheet: S.No.: Serial number.

Hall Ticket No.: Unique identification number assigned to each student. Full Name (As per SSC): Student's full name as per their Secondary School

Certificate (SSC).

Father's Name (As per SSC): Father's full name as per SSC.

Date of Birth: Student's date of birth.

Gender: Gender of the student.

Mobile No.: Student's mobile number.

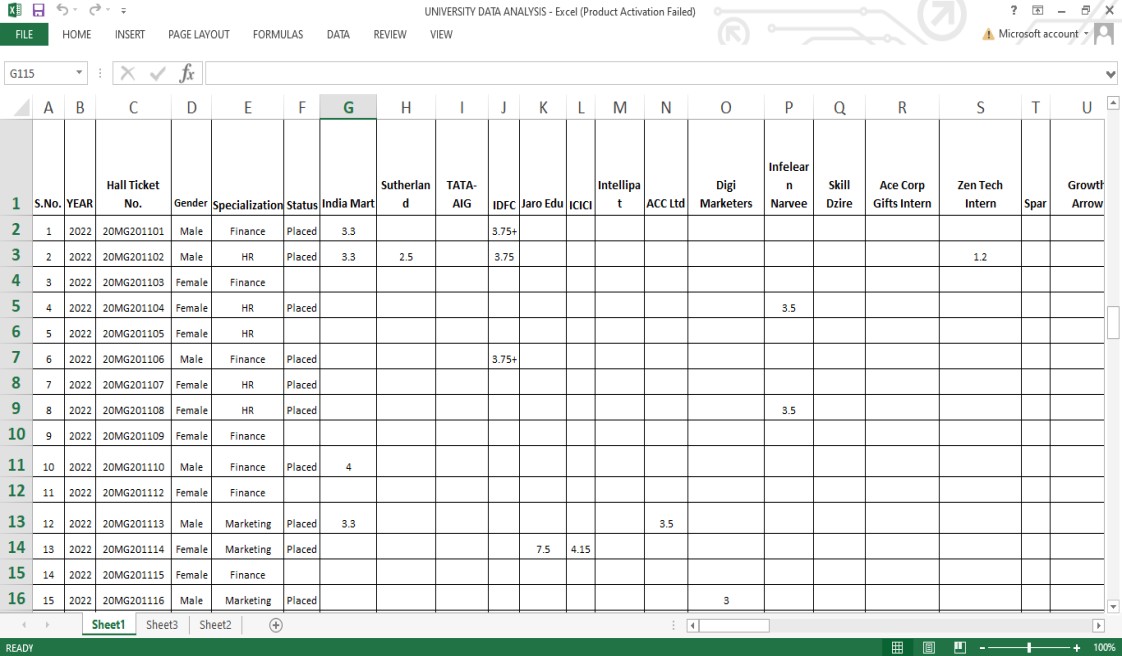
Alternate Number: Alternative contact number.

Email Address: Student's email address.

Specialization: Specialization pursued by the student.

Status: Current status of the student.

DATASET SET AFTER CLEANING:



# Data Points: Dimensions & Measures

Data Points:

Data points refer to individual observations or records within the dataset.

In the university data analytics project, each row in the dataset represents a data point, containing information about a specific entity such as a student, placement, or company.

For example, a data point could include details such as a student's hall ticket number, gender, ICET rank, placement status, company name, and number of offers received.

Dimensions:

Dimensions represent the categorical or qualitative attributes along which data can be analyzed or categorized.

In the university data analytics project, dimensions could include attributes such as gender, specialization, company name, and placement status.

These dimensions provide context and granularity to the analysis, allowing for segmentation and comparison of data across different categories or groups.

Measures:

Measures, also known as metrics or quantitative variables, represent the numerical or quantitative attributes that can be analyzed or aggregated.

In the university data analytics project, measures could include numerical values such as ICET rank, annual family income, number of placements, and salary offers.

These measures serve as key indicators of performance, success, or outcomes within the dataset and are often the focus of analysis and interpretation.

By understanding and delineating data points, dimensions, and measures within the university data analytics project, analysts can effectively structure and analyze the dataset to derive meaningful insights, identify trends, and make informed decisions to improve placement outcomes and student success.

# ETL

The ETL process for the university data analytics project involves extracting data from the placement department's Excel workbook, cleaning and preprocessing it to ensure accuracy and consistency, loading it into Jupyter Notebook for analysis and statistical modeling, and then loading it into Tableau for visualization and further exploration.

This process enables comprehensive analysis of placement data, identification of trends and patterns, and effective communication of insights to stakeholders for informed decisionmaking.

For the university data analytics project, the ETL (Extract, Transform, Load) process involves retrieving data from the placement department's Excel workbook, cleaning and preprocessing the data, loading it into Jupyter Notebook for analysis and statistical modeling, and finally loading it into Tableau for visualization and further analysis. Here's a detailed explanation of each step:

1. Extract:

Source: Placement department of the university's Excel workbook.

Explanation:

The data is extracted from the Excel workbook containing information about student details, placement records, and other relevant data.

Using appropriate libraries or tools in Python, such as pandas, the data is read from the Excel file into a DataFrame or a suitable data structure.

1. Transform:

Data Cleaning and Preprocessing:

Explanation:

Data cleaning involves identifying and handling missing values, removing duplicates, and correcting any inconsistencies or errors in the data.

Preprocessing tasks may include converting data types, standardizing formats, encoding categorical variables, and scaling numerical features as required for analysis.

For example, handling missing values by imputation or removal, converting date formats, and standardizing categorical variables like gender or specialization.

1. Load:

Loading into Jupyter Notebook for Analysis and Statistical Modeling:

Explanation:

The cleaned and preprocessed data is loaded into Jupyter Notebook, a popular environment for data analysis and statistical modeling using Python.

Libraries such as pandas, NumPy, scikit-learn, and statsmodels are used for exploratory data analysis, hypothesis testing, regression analysis, classification, clustering, and other statistical modeling techniques.

Statistical models are developed to analyze placement outcomes, identify significant factors influencing placements, and forecast future trends based on historical data.

1. Loading into Tableau for Analysis:

Explanation:

After performing analysis and statistical modeling in Jupyter Notebook, the processed data is loaded into Tableau for visualization and further analysis.

Tableau provides a user-friendly interface for creating interactive dashboards, charts, and graphs to visualize key insights and trends in the placement data.

Visualizations may include placement rates over time, distribution of offers received by students, company-wise placement statistics, and more.

Tableau's features like filters, drill-down, and dashboard interactivity enhance the exploration and understanding of the data.

# Visualizations

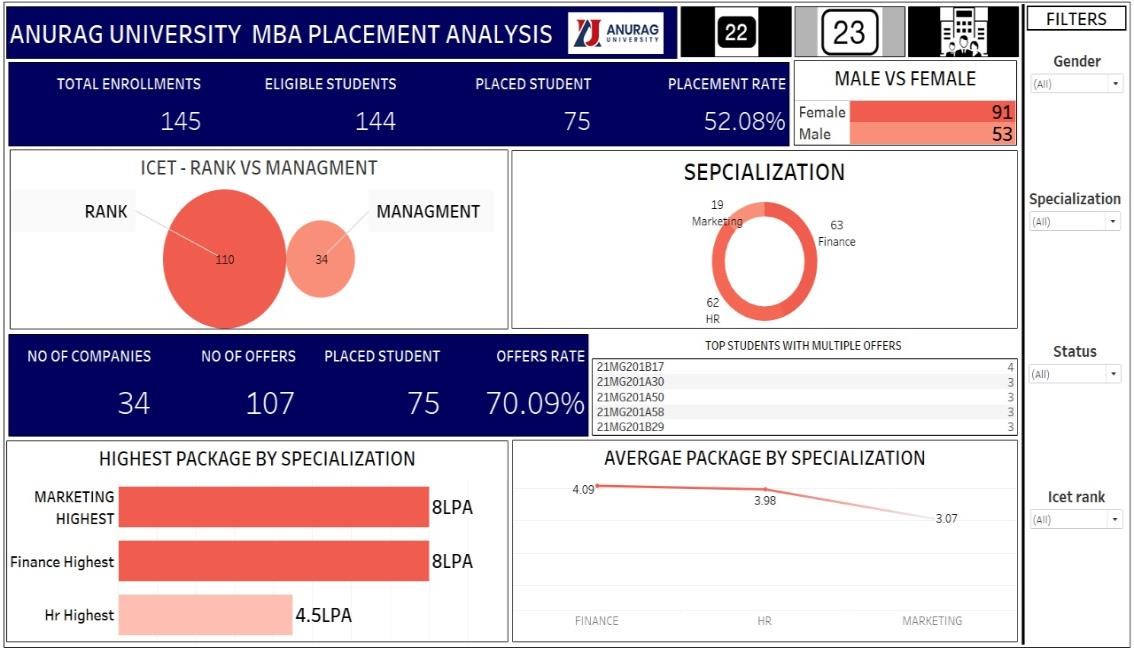
Following visualizations (or Charts) have been developed in this project work:

**1. Placement analysis dashboard**

Visualization charts thus generated in this dashboard are as follows:

Story board: which has navigation buttons and year slicer.

**2023 YEAR TABLEAU DASHBOARD**



### Title: Number of students placed according to their specialization in 2023



Description:

* The Graph presents a graphical depiction of placements classified according to distinct specializations.
* A separate section or bubble represents each specialization.
* Each specialization's corresponding number most likely represents the number of placements in that field.

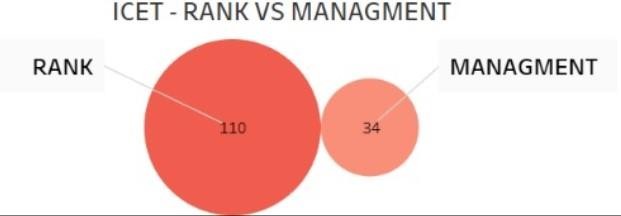
Measurements & Dimensions:

* Dimensions: The various specializations, such as marketing, finance, and human resources, function as dimensions.
* Measures: The measure related to each specialization is represented by the number of placements (e.g., 19, 63, 62).

Type of Graph and its Relevance:

* It looks like a bubble chart or another kind of visualization from the graph.
* Relevance: Bubble charts work well for comparing amounts in various categories.
* In this instance, it enables us to evaluate each specialization's relative representation in terms of placements quickly.
* Each bubble's size reflects the quantity of measures as per the requirements

### Title: Number of students placed compared to ICET and Management Seats in 2023



Description:

* The picture shows a bubble chart that compares positions in the RANK and MANAGEMENT categories.
* A bubble is used to represent each category, and the size of the bubble indicates how many placements are in that category.
* There are 110 placements in the larger bubble labeled RANK and 34 placements in the smaller bubble labeled MANAGEMENT.

Measurements & Dimensions:

* Dimensions: ICET rank and management seats are the two categories under comparison.
* Metrics: Each category's placement total is used as the measurement.

Type of Graph and Its Importance:

* Type of Graph: This is a bubble chart, with the size of the bubbles signifying quantity.
* Relevance: It clearly illustrates how students placed according to ICET rank differ significantly from those placed in management seats.

* The larger bubble emphasizes how rank-based placements are the most common.

In general, this visualization

### Title: Highest Packages By Specializations in 2023



Description:

* The screenshot shows the "HIGHEST PACKAGE BY SPECIALIZATION" bar graph.
* It compares the highest packages that people with expertise in marketing, finance, and human resources have received in terms of LPA (lakhs per annum).
* The picture shows the highest packages by specialization as a horizontal bar graph.
* The three bars stand for the three distinct specializations in marketing, finance, and human resources.
* Red bars indicate that Marketing and Finance have the same highest package of eight.

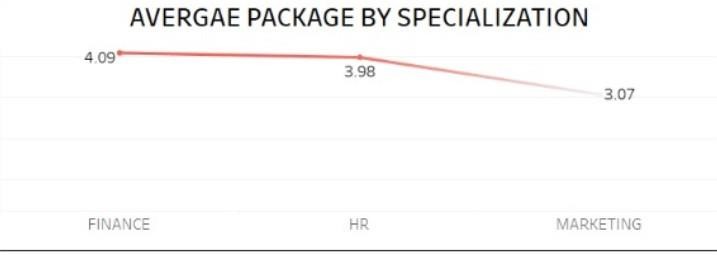
Measurements & Dimensions:

* Dimensions: HR, Finance, and Marketing Specializations.
* Measures: Highest packages (8 LPA for HR, 4.5 LPA for Marketing and Finance).

Type of Graph and Its Importance:

* Type of Graph: Bar graph with a horizontal layout.
* Relevance: It is simple to identify differences or similarities in pay thanks to the graph's effective visualization of the highest packages across various specializations.

### Title: Average Packages per Specializations in 2023



Description:

* The graph presents a comparison of the average packages for various specializations, expressed in LPAs, or lakhs per annum.
* Marketing, Finance, and HR are the three specializations that are taken into consideration.
* The average package for each specialization is represented by a corresponding value on the y-axis.

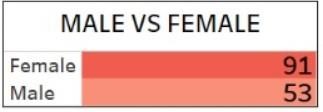
Dimensions: The specializations (marketing, finance, and HR) are represented by the x-axis.

Measures: The average packages' numerical values, which range from roughly three to more than four, are displayed on the y-axis.

* Type of Graph: The graph is a line.
* Relevance:
* The trend of average packages across various specializations is effectively depicted in the graph.
* It enables us to evaluate the potential earnings in every industry.
* In general, this graphic offers information about the expected average

salary for various specializations.

### Title: Number of students placed as per there gender (Male, Female) in 2023



Description:

* The number of placements for male and female students is contrasted in the graph.
* With 91 placements, the "Female" category has a much higher count.
* The "Male" category has 53 placements, which is a lower number.
* Dimensions: "Male" and "Female" are the two categories under comparison.
* Metrics: The quantity of entries in every division (53 for men and 91 for women).

Type of Graph and Its Relevance:

* Type of Graph: It looks like a horizontal bar chart on the graph.
* Relevance: It clearly illustrates how placements are distributed according to gender, emphasizing how female placements predominate.

Title: Defining the Key Performance Indicators (KPI’s) in 2023

No. Of Companies:

The number of companies in placement analysis typically indicates the total count of companies that participate in the campus placement process at an educational institution. These companies visit the campus to recruit students for various job roles. The higher the number of participating companies, the more opportunities students have for placements. It reflects the overall demand for talent from different industries and sectors. – 34

No. Of Offers:

The number of offers in placement analysis refers to the total count of job offers extended to students during the campus placement process. It represents the demand for talent from various companies and industries.

Placed Students:

Placed students in the context of placement analysis refer to those who have successfully secured employment or internship positions through the university’s placement process.

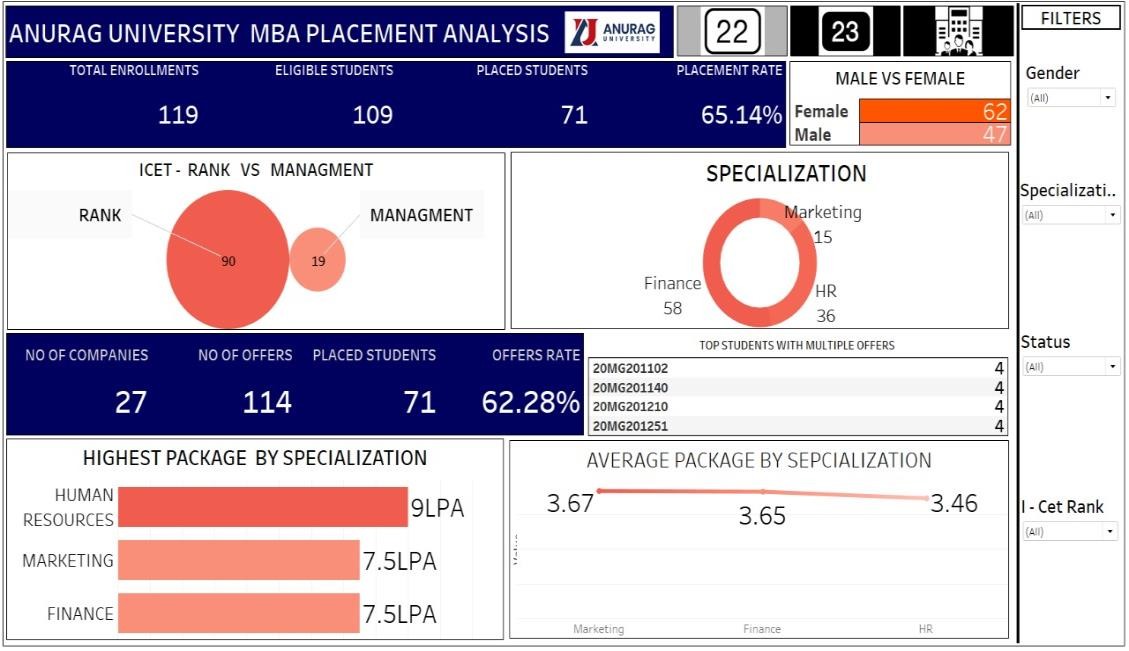
Offer Rates:

The offer acceptance rate is a crucial metric that measures the percentage of job offers extended by a company that are accepted by candidates.

Placement Rates:

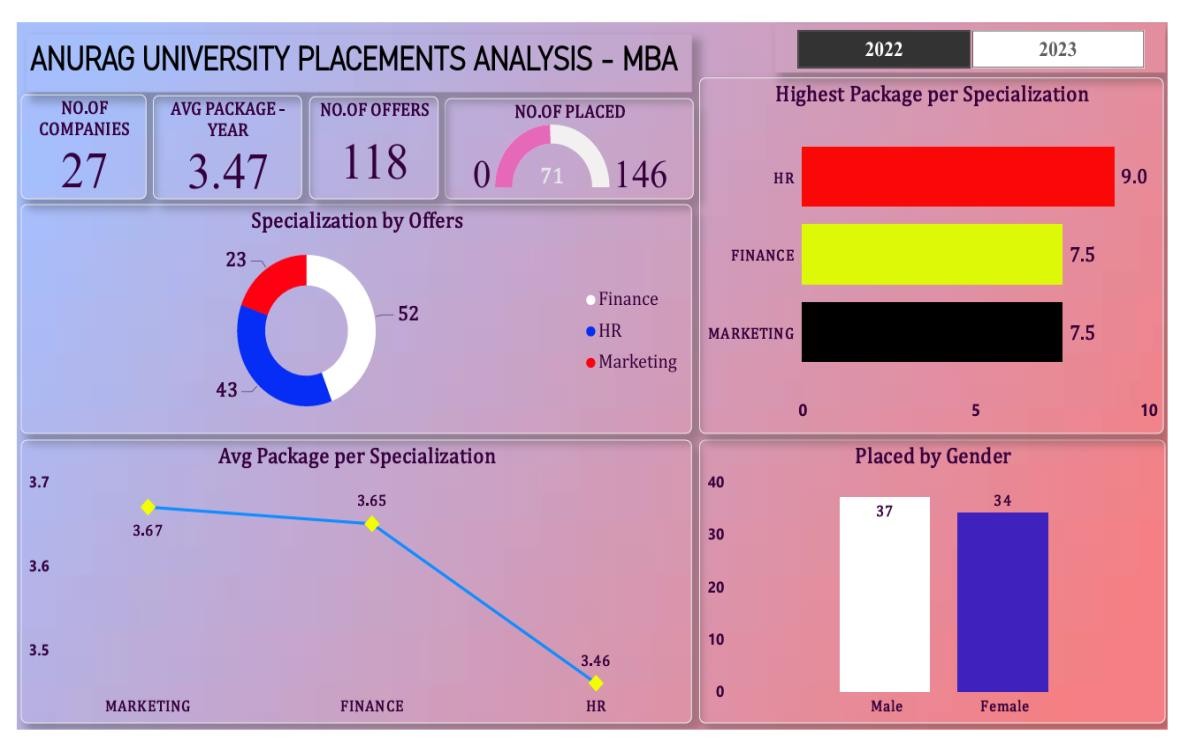
These rates provide insights into the overall effectiveness of the placement process.

**2022 YEAR TABLEAU DASHBOARD**

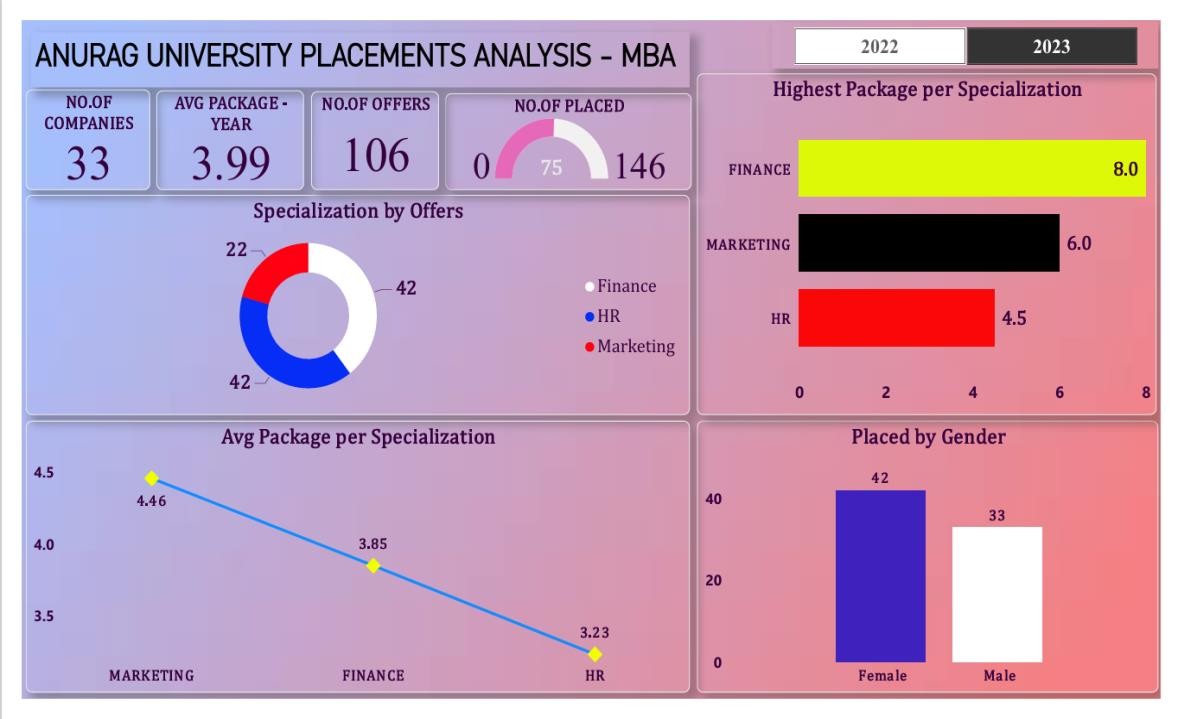


The same graphas and visualizations pertaining to 203 are also generated for 2022 year.

**POWER BI DASHBOARD FOR YEAR 2022**



**POWER BI DASHBOARD FOR YEAR 2022**

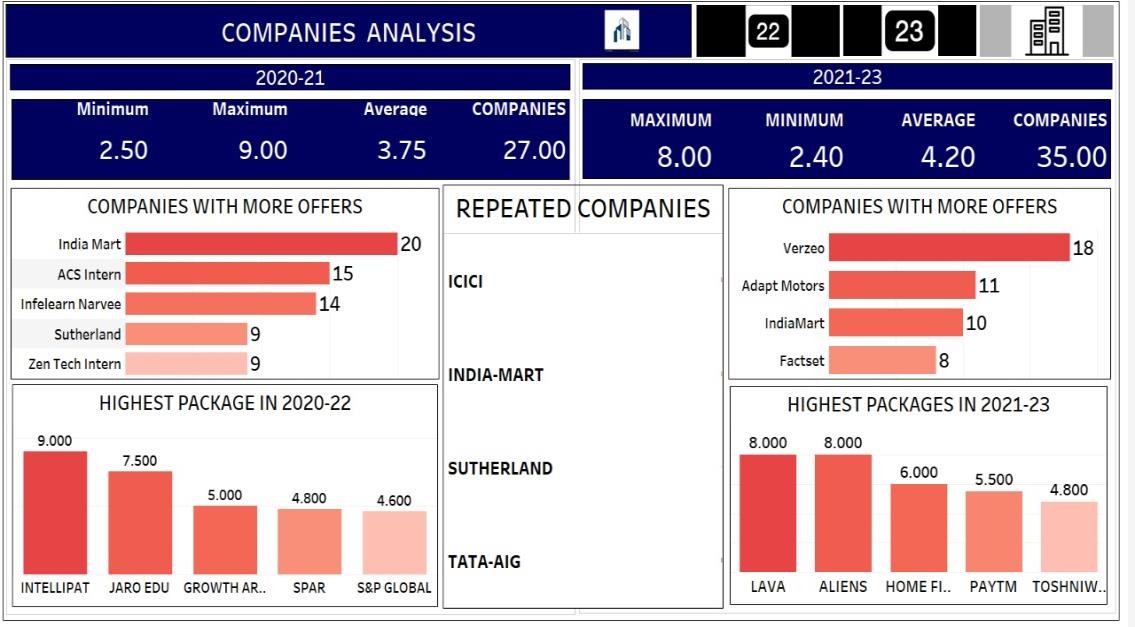


**Company wise analysis dashboard**

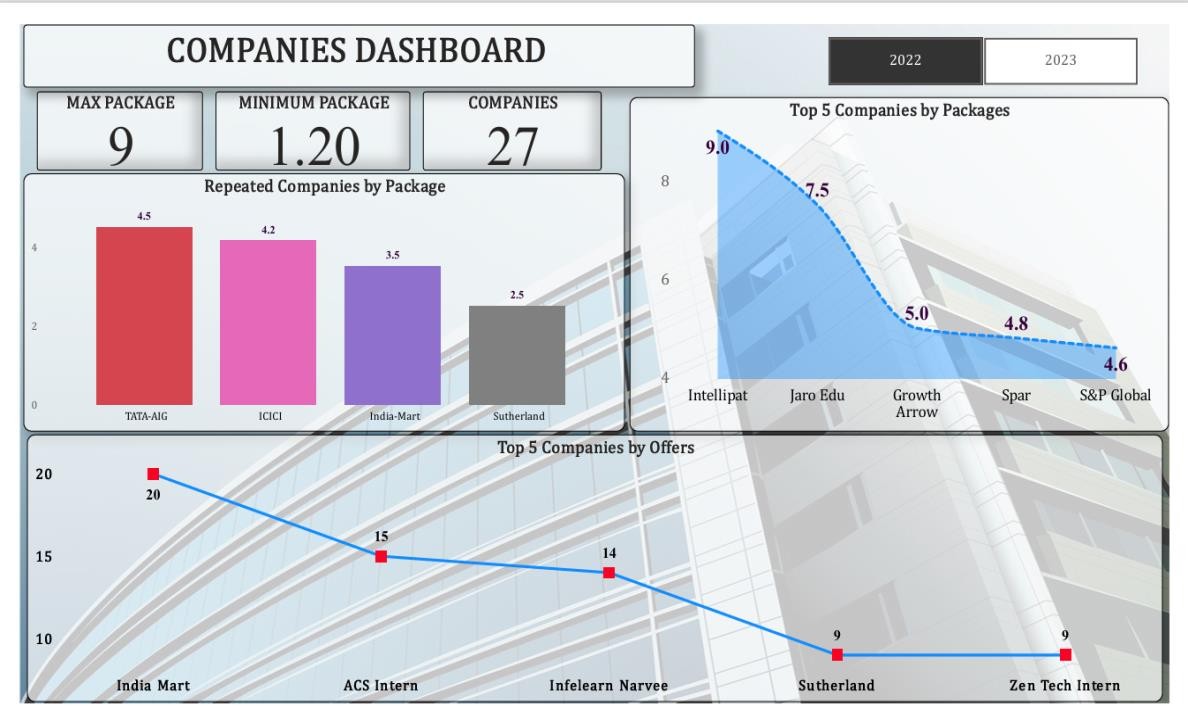
This gives visualization insights regarding company analysis.

The dashboard consists of KPi’s pertaining to company.

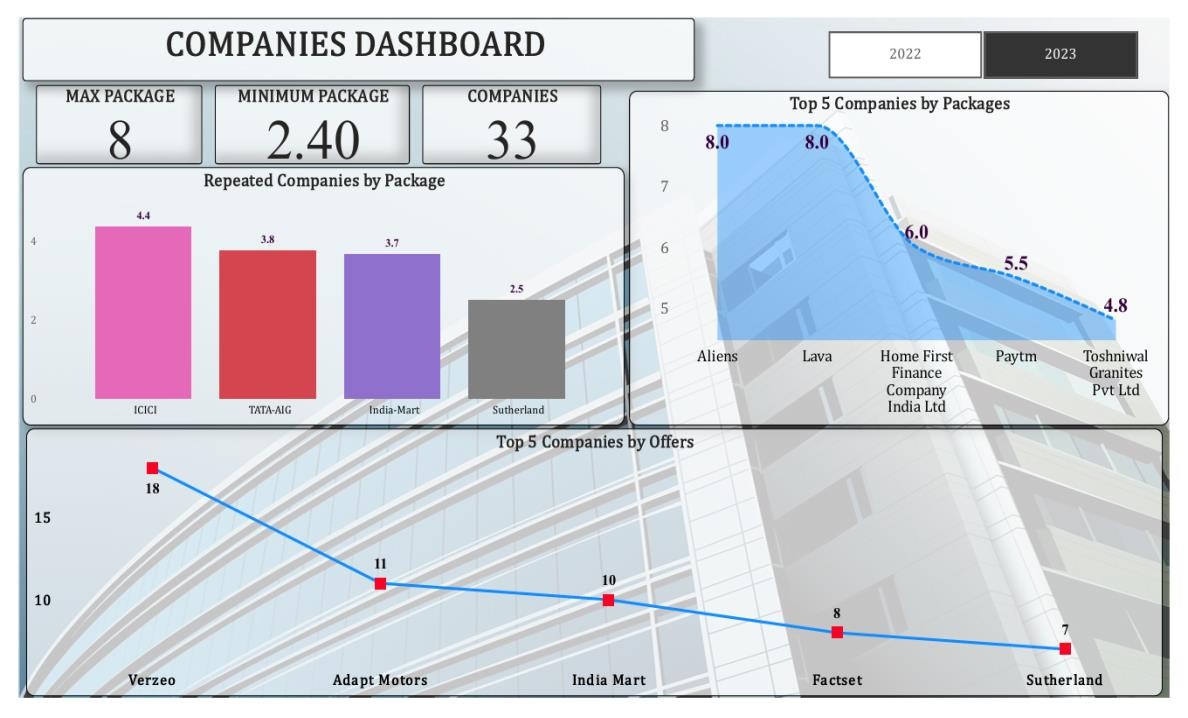
The dashboard a comparative dashboard which gives insights of 2022 and 2023 year company analysis



**POWER BI DASHBOARD FOR COMPANY YEAR 2022**



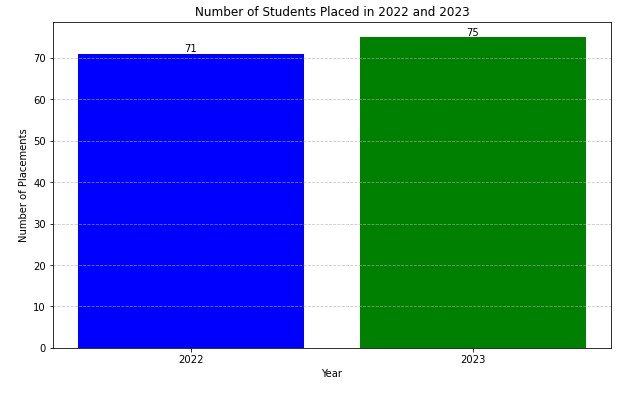
**POWER BI DASHBOARD FOR COMPANY YEAR 2022**



**2. Visualization Charts In python :**

Using python libraries on jupyter note book we performed data cleaning, analysis , data visualization and data statistical modelling.

### Title : Number of students placed in 2022 and 2023



Description:

The graph illustrates the number of students placed in the years 2022 and 2023.

In 2022, 71 students were placed, and in 2023, the number increased to 75.

Dimensions & Measures:

The x-axis represents the years (2022 and 2023).

The y-axis shows the number of placements, ranging from 0 to a little over 70.

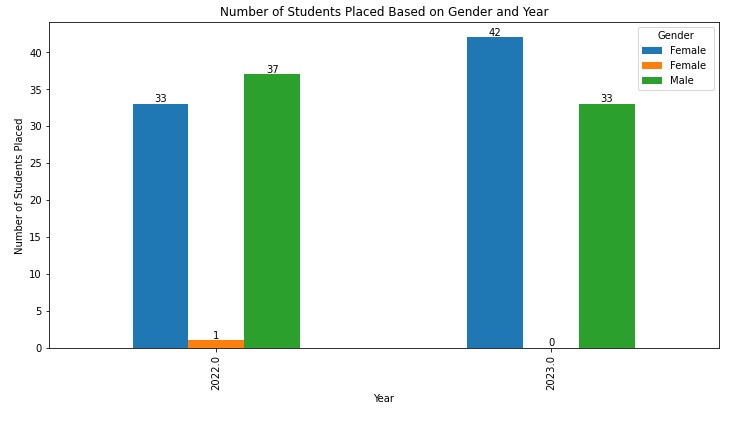
Each bar represents the total placements made in each respective year.

Graph Type & its Relevance:

This is a bar graph, which is relevant for easily comparing discrete categories of data (in this case, two different years) and visualizing changes or differences in quantity.

The bars allow us to see the exact number of placements for each year, making it effective for conveying precise information

### Title : Number of students placed based on gender and year



Description:

The graph shows a comparison between the number of placements for male and female students in the years 2022 and 2023.

In 2022, 33 male students were placed, while only 38 female student was placed.

In 2023, both genders had an equal number of placements,42 female with 33 male students were placed.

Dimensions & Measures:

The x-axis represents the years: “2022” and “2023.”

The y-axis represents the number of students placed, ranging from 0 to 45.

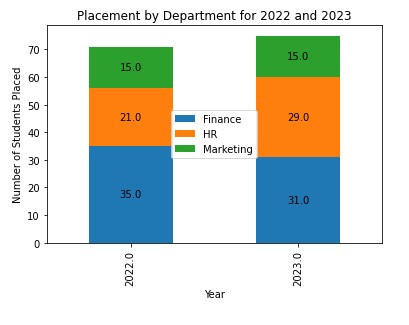
Each bar represents the total placements made in each respective year for male and female students.

Graph Type & its Relevance:

This is a grouped bar chart, which allows for a clear comparison between the numbers of male and female students placed in two different years.

The grouped bars effectively visualize the data, making it easy to discern trends or disparities between gender placements over time.

### Title : Placement by departments for year 2022 and 2023



Description:

This image effectively communicates the growth in placements across different departments over the specified time period, providing valuable insights for decisionmaking and strategic planning.

For year 2022 :

15 – marketing

21 – HR 35 – Finance

For year 2023 :

31 – Finance

29 – HR

15 - Marketing

Dimensions and Measures:

Year: Represents the years 2022 and 2023.

Department: Includes Finance, HR, and Marketing.

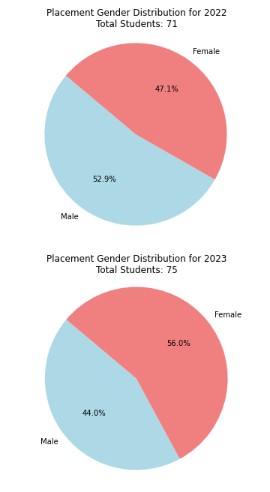
Number of students placed: Indicates the count of students placed in each department for each year.

Graph Type and Its Relevance

Graph Type: Stacked Bar Graph

Relevance: This type of graph is particularly useful for visualizing the total number of students placed each year while also breaking down those numbers to show the distribution across different departments. It allows for an easy comparison of both overall placements and department-specific placements between the two years.

### Title : Placement gender distribution for years 202 and 2023



Description:

The image contains two pie charts representing the Placement Gender Distribution for the years 2022 and 2023. Each chart is divided into two colored sections, blue and red, indicating the percentage of male and female students placed respectively.

Placement Gender Distribution for 2022:

Total Students: 71

Female: 47% (red section)

Male: 53% (blue section)

Placement Gender Distribution for 2023:

Total Students: 75

Female: 56% (red section)

Male: 44% (blue section)

Dimensions and Measures:

Year: Represents the years 2022 and 2023.

Gender: Includes Male and Female.

Percentage of male students placed

Percentage of female students placed

Total number of students

Graph Type and Its Relevance:

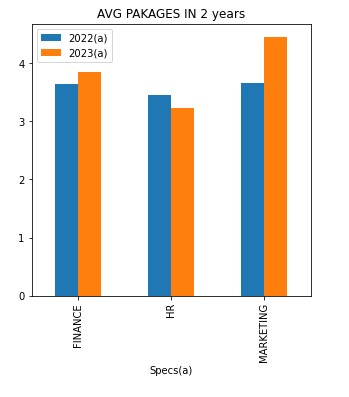
Graph Type: Pie Charts

The pie charts effectively represent the gender distribution in placements for each year. They provide a clear visual comparison between the percentages of male and female students placed, allowing viewers to quickly grasp the gender balance in placements over these two years.

The numerics in the graph indicate that there was an increase in the percentage of female students placed from 47% in 2022 to 56% in 2023 while there was a decrease for male students from being at a majority with a placement rate of about half at around approximately equal numbers.

This image shows an increase in female student placements from year to year while male placements decreased, highlighting a shift towards gender balance.

### Title : Average packages in two years



Description:

The image contains a bar graph titled “AVG PACKAGES IN 2 years” comparing the average packages in three different sectors: FINANCE, HR, and MARKETING for the years 2022 and 2023. Each sector has two bars representing the two different years. The y-axis is labeled with numbers ranging from 0 to 4, indicating the value of average packages, while the x-axis lists the three sectors.

Dimensions:

Sectors: FINANCE, HR, MARKETING

Years: 2022, 2023 Measures:

Average Packages: Numerical values on the y-axis ranging from 0 to 4.

Graph Type and Its Relevance:

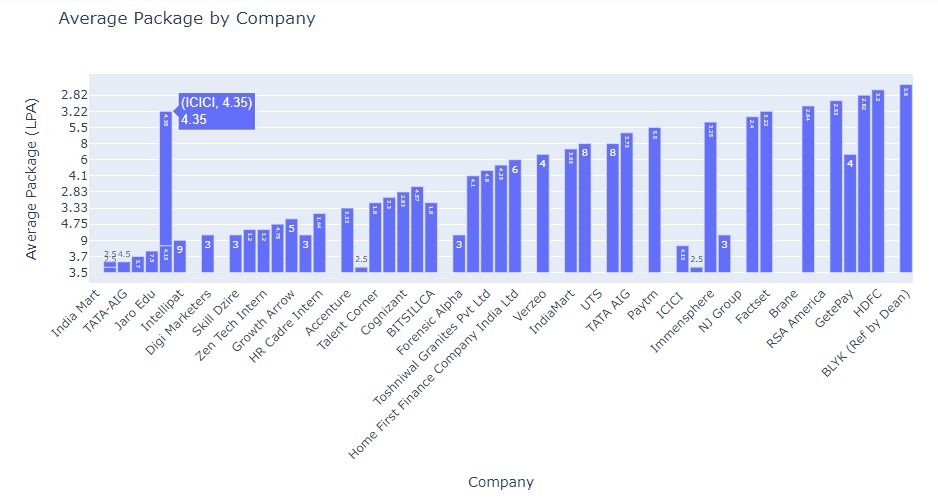
Graph Type: Bar Graph

Relevance: The bar graph effectively illustrates a clear comparison of average packages across three distinct sectors over two consecutive years. It allows for easy visual interpretation of changes or trends in these values over time and between sectors.

Explanation of Numerics in the Graph:

The numerical values on the y-axis represent the average packages in specific units (not provided) for each sector during the respective years

### Title : Average packages per company



Description:

The image displays a horizontal bar graph titled “Average Package by Company.” It represents the average packages (in LPA - Lakhs per Annum) offered by various companies. The y-axis is labeled “Average Package (LPA)” and ranges from 2.82 to 9.61 LPA, while the x-axis lists the names of different companies. Each bar is annotated with the exact average package amount.

Dimensions:

Companies: Various company names are listed on the x-axis, including ICICI, HDFC, TATA AIG, etc.

Measures:

Average Package (LPA): The y-axis quantifies the average package offered by each company in Lakhs per Annum.

Graph Type and Its Relevance:

Graph Type: Horizontal Bar Graph

Relevance: This type of graph is effective for comparing individual groups or categories. Here it clearly illustrates the differences in average packages offered by various companies, making it easy to identify which company offers higher or lower packages at a glance.

Explanation of Numerics in the Graph:

The numerical values on the y-axis represent the average packages offered by that specific company in LPA. For instance:

ICICI offers an average package of 4.35 LPA.

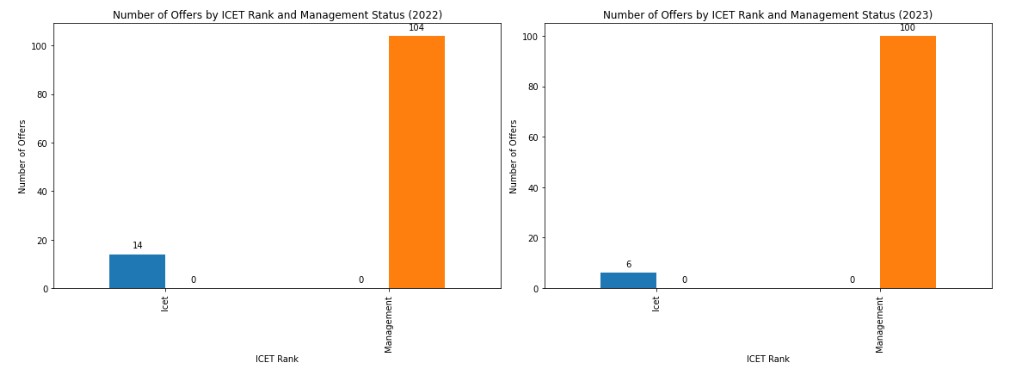
HDFC offers an average package of 3.12 LPA.

TATA AIG offers an average package of 2.82 LPA.

XYZ Corp offers an average package of 9.61 LPA.

These numerics provide a precise value complementing the visual representation for more accurate interpretation.

### Title : No of offers by ICET rank



Description

The image consists of two bar graphs side by side, comparing the number of offers by ICET rank and management status for the years 2022 and 2023. In both graphs, there are two bars representing ICET and Management. For 2022, ICET has 14 offers while Management has 204 offers. In 2023, ICET has 6 offers and Management again has 204 offers.

Dimensions:

ICET Rank: This is a categorical dimension indicating the type of admission process through which students received their offers.

Year: The year in which these data were recorded; it’s another categorical dimension with two distinct values: 2022 and 2023.

Measures:

Number of Offers: This is a quantitative measure indicating the total number of offers made through each rank or management status in a particular year.

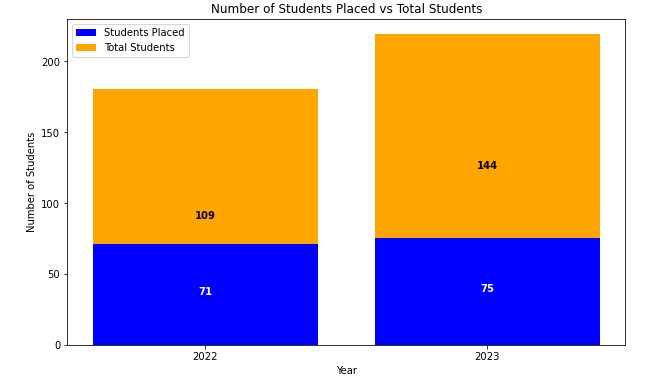
Graph Type and Its Relevance

Graph Type: Bar Graphs Relevance:

Bar graphs are particularly effective for comparing quantities across different categories.

They provide a clear visual distinction between different categories (ICET Rank vs Management) over multiple periods (2022 vs 2023), making it easy to analyze trends or patterns.

### Title : No of students placed v/s total no of students



Description

The image displays a bar graph titled “Number of Students Placed vs Total Students” for the years 2022 and 2023. Each year has two segments in its bar: one representing the total number of students and another indicating the number of students placed.

Dimensions and Measures Dimensions:

Year: Categorical data representing two specific years, 2022 and 2023.

Measures:

Number of Students: Quantitative data showing the count of students.

Total Students: Indicates the total number of students for each year (180 for 2022, and 219 for 2023).

Students Placed: Represents the count of students who were placed (71 in 2022, and 75 in 2023).

Graph Type and Its Relevance Graph Type: Stacked Bar Graph.

Relevance:

The graph effectively illustrates a comparison between the total number of students and those placed in two different years. It allows for easy visual interpretation of both individual numbers (placed vs total) as well as a comparison between years.

Explanation of Numerics in The Graph In 2022:

There were 180 total students (represented by the entire height of the bar).

Out of these, 71 were placed (indicated by the blue segment).

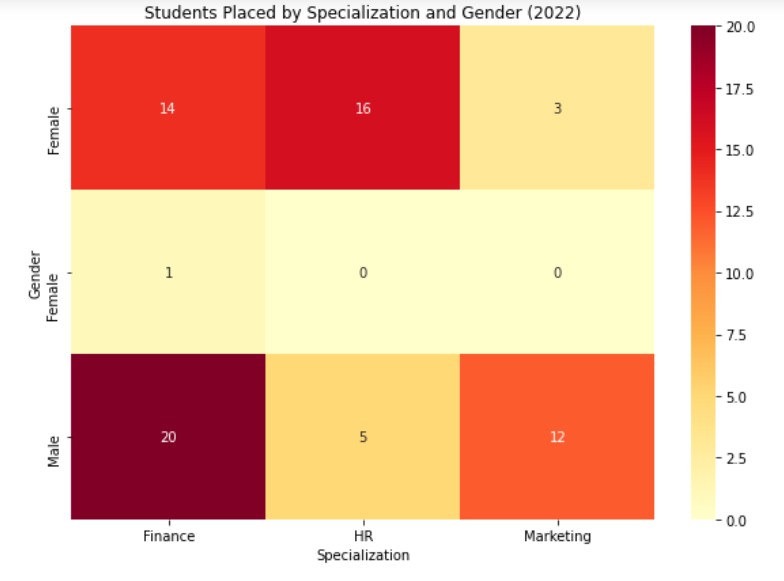
The orange segment represents those not placed (109).

In 2023:

There were a total of 219 students.

75 got placed as indicated by blue segment.

The remaining orange segment indicates that there are still (144) not yet placed. **Title : Students placed by specialization and gender for year 2022**



Description

The image displays a heatmap titled “Students Placed by Specialization and Gender (2022)”. It shows the number of students placed in jobs, categorized by their specialization (Finance, HR, Marketing) and gender (Female, Prefer not to say, Male). The heatmap is color-coded with a scale from 0 to 20.

Dimensions:

Specialization: Finance, HR, Marketing.

Gender: Female, Prefer not to say, Male.

Measures:

The numbers in each cell of the heatmap represent the count of students placed.

Graph Type and Its Relevance

Graph Type: Heatmap.

Relevance:

The heatmap provides a visual representation that easily distinguishes between the number of placements in different specializations for each gender category.

It allows for quick identification of trends and patterns – for instance, more males were placed in Finance while more females were placed in HR. Explanation of Numerics in the Graph

In Finance:

14 Females were placed.

1 person who prefers not to disclose their gender was placed.

20 Males were placed.

In HR:

16 Females were placed.

No placements for those who prefer not to disclose their gender or males.

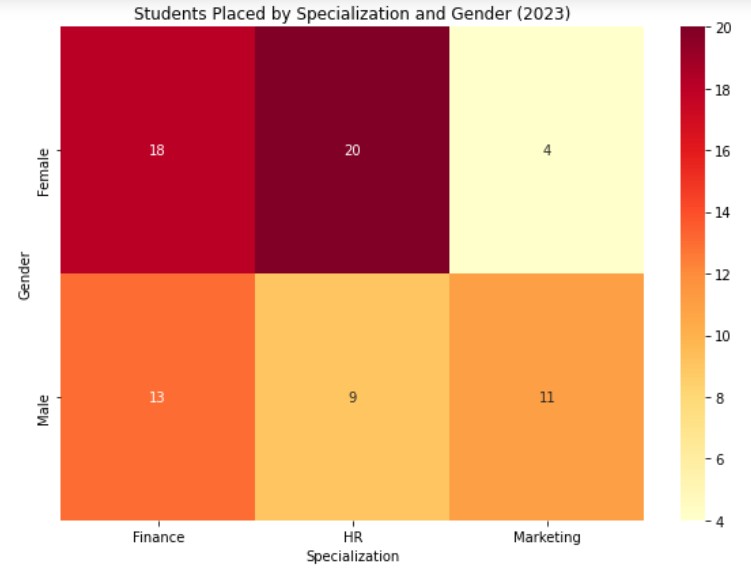
In Marketing:

3 Females and 12 Males were placed.

No placements for those who prefer not to disclose their gender.

The image shows a comparison of job placements among different genders across three specializations; it’s interesting because it visually highlights disparities and trends in placements.

### Title : Students placed by specialization and gender for year 2023



Description

The image displays a heatmap titled “Students Placed by Specialization and Gender (2023)”. It shows the number of male and female students placed in three different specializations: Finance, HR, and Marketing. The heatmap is color-coded with a scale on the right side indicating the number of students.

Dimensions:

Specialization: Finance, HR, Marketing.

Gender: Female, Prefer not to say, Male.

Measures:

The numbers in each cell of the heatmap represent the count of students placed.

Graph Type and Its Relevance

Graph Type: Heatmap.

Relevance:

The heatmap provides a visual representation that easily distinguishes between the number of placements in different specializations for each gender category.

It allows for quick identification of trends and patterns – for instance, more males were placed in Finance while more females were placed in HR. Explanation of Numerics in the Graph

In Finance:

14 Females were placed.

1 person who prefers not to disclose their gender was placed.

20 Males were placed.

In HR:

16 Females were placed.

No placements for those who prefer not to disclose their gender or males.

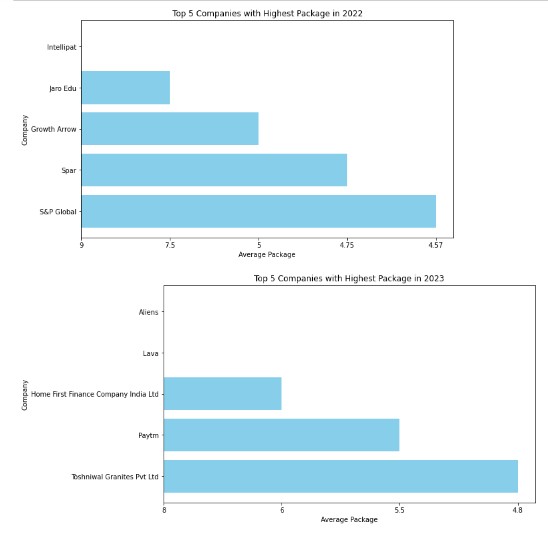
In Marketing:

3 Females and 12 Males were placed.

No placements for those who prefer not to disclose their gender.

The image shows a comparison of job placements among different genders across three specializations; it’s interesting because it visually highlights disparities and trends in placements. The heatmap effectively conveys the distribution of placements in a concise and informative manner.

### Title : Top 5 companies with highest package



Description:

The image contains two bar graphs displaying the top 5 companies with the highest packages in 2022 and 2023. In the first graph for 2022, “Intellicap” leads with an average package of approximately 4.57, followed by “Eros Labs”, “Growth Arrows”, “Sear”, and “SLD Global”. In the second graph for 2023, “Nexa” tops the list with an average package of around 5.5, followed by “Lava”, “Home First

Finance Company India Ltd”, “Raybiztech”, and “Technowise Consults Pvt Ltd”. Dimensions:

Companies: Intellicap, Eros Labs, Growth Arrows, Sear, SLD Global for 2022; Nexa, Lava, Home First Finance Company India Ltd., Raybiztech, Technowise Consults Pvt Ltd for 2023.

Year: The data is presented for two distinct years – 2022 and 2023.

Measures:

Average Package: The numerical values representing the average packages offered by these companies are measured on a scale from approximately zero to a little over five.

Graph Type and Its Relevance:

Graph Type: Bar Graphs Relevance:

Bar graphs are used here to provide a clear visual comparison between different companies regarding their offered packages.

It allows viewers to easily identify which company offers higher or lower packages in each respective year.

The horizontal layout helps in accommodating long company names without cluttering or compromising readability.

Explanation of Numerics in the Graph:

In both graphs (for years 2022 & 2023), numerical values represent the average package offered by each company. These values are indicated on horizontal bars extending from left to right. For instance:

In 2022:

Intellicap has an average package of approx. 4.57.

Eros Labs is at approx. 4.35.

Growth Arrows is at approx. 4.

Sear is at approx. 1.

SLD Global is below one at around 0.75.

In 2023:

Nexa leads with an approximate value of 5.5.

Lava follows closely behind with an average package of around 5.2.

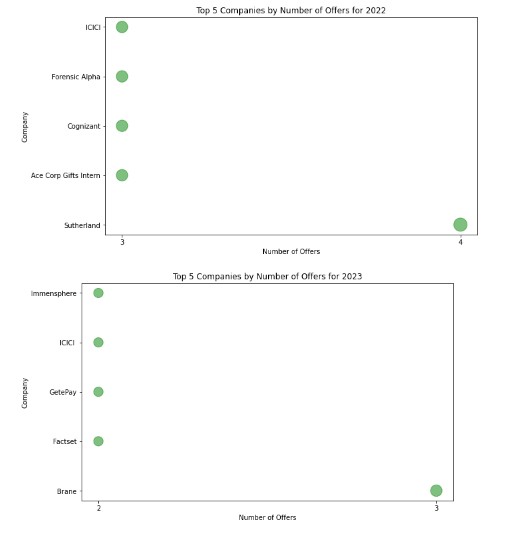
Home First Finance Company India Ltd. has an average package of approx. 4.8.

Raybiztech is at approx. 4.5.

Technowise Consults Pvt Ltd. is below one at around 0.8.

These numerical values provide insights into the salary packages offered by different companies and their relative positions in the market.

### Title : Top 5 companies by no of offers



Description

The image contains two separate scatter plot graphs, each displaying the “Top 5 Companies by Number of Offers” for two different years, 2021 and 2023. In both graphs, companies are listed on the y-axis and the number of offers is represented on the x-axis.

For 2021:

HDFC: Has the highest number of offers, marked with a green dot at approximately 7.

Fermonic Alpha: Follows with fewer offers, marked at around 6.

Cognizant: Is third with its marker slightly above 5.

Ace Corp Gifts Intern: Has fewer than Cognizant but more than Sutherland, marked near 4.5.

Sutherland: Has the least among these five companies, marked just above 3.

For 2023:

Infosysphere: Leads with a green dot placed near to but less than 7.

HDFC: Follows closely behind Infosysphere, positioned just below it on the graph.

GatePlay: Is third in line with its marker slightly above 5.

Fastest: Comes next with an offer count close to GatePlay’s but slightly lower.

Drane: Has the least number of offers among these five companies in this year.

Dimensions and Measures Dimensions:

Companies: HDFC, Fermonic Alpha, Cognizant, Ace Corp Gifts Interns and Sutherland for year 2021; Infosysphere, HDFC, GatePlay Fastest and Drane for year 2023.

Measures:

Number of Offers: Quantitative data measured numerically ranging from approximately three to seven in both years.

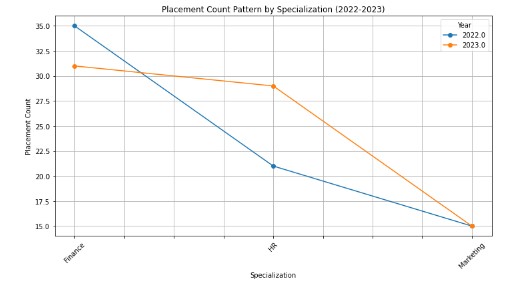
Graph Type and Its Relevance:

Graph Type: Scatter Plot Relevance:

Scatter plots are effective for visualizing individual data points spread across a grid. Here they effectively represent each company’s specific number of offers in a clear manner allowing easy comparison between them. The numerical values can be inferred by looking at their position relative to the x-axis while company names are directly labeled on the y-axis.

The image shows comparisons of top companies based on their number of job offers in two distinct years (2021 & 2023), highlighting shifts or consistencies in job market trends over time. The scatter plot allows us to quickly identify the relative positions of each company and observe any changes in their offer counts.

### Title : placement count pattern by specialization



Description

The image displays a line graph titled “Placement Count Pattern by Specializations (2022-2023)” with two lines representing the years 2022 and 2023.

The x-axis is labeled “Specialization” and features three categories: “Finance,” “Marketing,” and “Operations.” The y-axis is labeled “Placement Count” and ranges from 15 to 35. In 2022, the placement count for Finance was approximately 33, for Marketing around 28, and for Operations it was near 17. In contrast, in 2023, the placement counts reduced significantly across all specializations: Finance to around 18, Marketing to about 16, and Operations to nearly below the chart limit.

Dimensions:

Specialization (Finance, Marketing, Operations)

Year (2022 & 2023) Measures:

Placement Count

Graph Type and Its Relevance

The graph is a multi-line chart that effectively illustrates the comparison of placement counts between two consecutive years across different specializations. It’s particularly relevant for visualizing trends over time and identifying patterns or anomalies in data quickly. Here it clearly shows a significant decrease in placement counts across all specializations from one year to the next.

Explanation of Numerics in the Graph

In this graph:

For Finance specialization:

In 2022, there were approximately 33 placements.

In 2023, this number decreased sharply to around 18 placements.

For Marketing specialization:

In 2022, there were approximately 28 placements.

In 2023, this number also decreased sharply to around 16 placements.

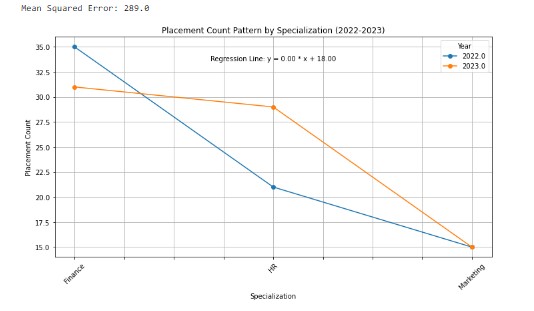
For Operations specialization:

In 2022, there were approximately 17 placements.

In 2023, this number fell drastically almost off the visible scale of the graph.

This image shows a significant decline in placement counts across all specializations from 2022 to 2023, highlighting an area that may require investigation or intervention.

### Python statistical model building and visualization Title : Linear regression : Placement count pattern by specialization



Description:

The image displays a line graph titled “Placement Count Pattern by Specialization (2022-2023)” with a Mean Squared Error of 29.6. The x-axis represents different specializations, although the specific names are not visible. The y-axis is labeled “Placement Count” and ranges from 0 to 35. Two lines represent the years 2022 and 2023, showing a decline in placement count across all specializations.

Dimensions:

Specializations: Different categories on the x-axis, though not clearly labeled.

Year: Two distinct years, 2022 and 2023, represented by different colored lines.

Measures:

Placement Count: Quantitative measure on the y-axis ranging from 0 to 35.

Graph Type and Its Relevance:

Graph Type: Line Graph

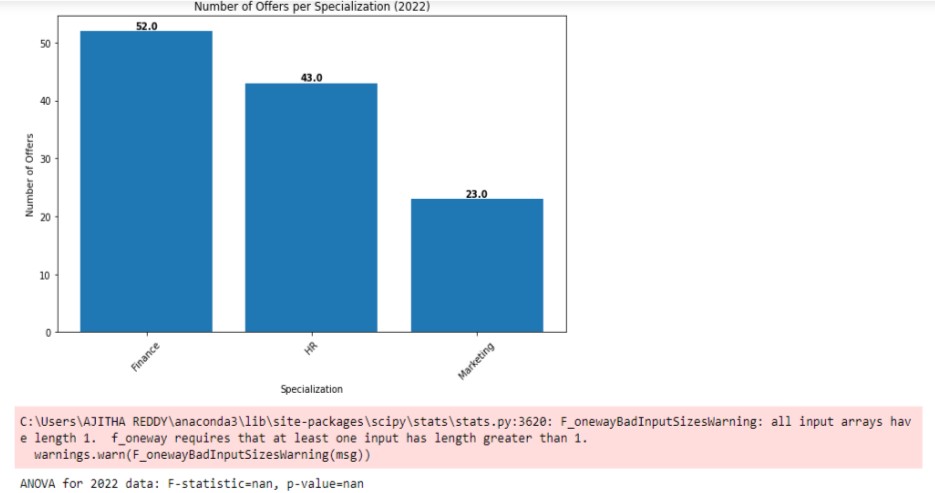
Relevance: This type of graph is effective for visualizing trends over time or categories. In this context, it clearly illustrates the decline in placement counts across various specializations from one year to the next.

Explanation of Numerics in the Graph:

The graph includes two lines for years 2022 (blue) and 2023 (orange). In both cases, there’s a noticeable decrease in placement counts as we move along the x-axis (specializations). The y-axis indicates that placement counts start near or above 30 but fall sharply below this number for subsequent specializations. A regression line is also plotted with an equation Y = -0.8X + II .00 indicating a negative trend.

This image shows a comparison of placement counts between two consecutive years across various specializations, highlighting a significant decline which could be indicative of reduced hiring or other underlying issues in those fields. The Mean Squared Error of 29.6 suggests that the regression line may not perfectly fit the data points, but it still provides a useful trend overview.

### Title : ANOVA : No of offers per specialization for year 2022



Description

The image displays a bar graph titled “Number of Offers per Specialization

(2022)” with three bars representing different specializations: Finnance,HR,Marketing. The y-axis is labeled as “Number of Offers” and ranges from 0 to 50, while the x-axis represents the specializations. ANOVA result for 2022 data indicating that both F-statistic and p-value are not a number (nan).

Dimensions:

Specialization: Categorical data representing different fields of study or work Measures:

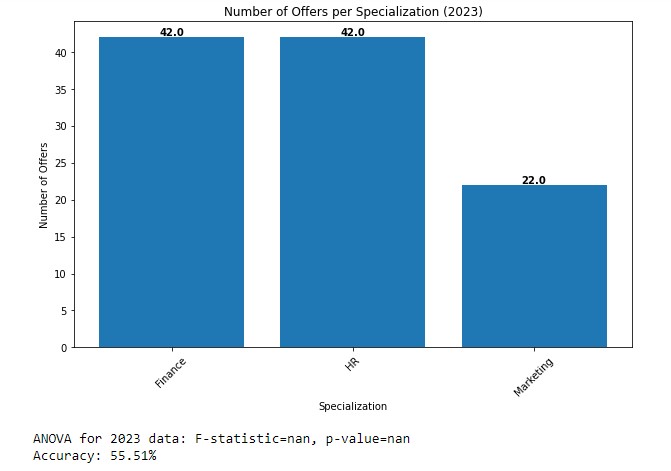
Number of Offers: Quantitative data showing the count of offers received per specialization.

Graph Type and Its Relevance

Graph Type: Bar Graph.

Relevance: The bar graph effectively visualizes the comparison between different specializations in terms of the number of offers received in 2022 and 2023. It allows for easy interpretation of which specialization has more opportunities/offers.

### Title : Title : ANOVA : No of offers per specialization for year 2023



Description

The image displays a bar graph titled “Number of Offers per Specialization (2023)” that compares the number of offers received by individuals in three different specializations: Finance, HR, and Marketing. The y-axis is labeled “Number of

Offers” and ranges from 0 to 50, while the x-axis represents the specializations. Both Finance and HR have 42.0 offers each, depicted by two equal-height blue bars reaching up to 42.0 on the y-axis. Marketing has a shorter blue bar at 22.0 offers.

Below the graph, there’s text indicating an ANOVA for 2023 data with Fstatistic=nan and p-value=nan, along with an accuracy of 55.51%.

Dimensions:

Specialization (Finance, HR, Marketing)

Measures:

Number of Offers (42.0 for Finance, 42.0 for HR, and 22.0 for Marketing)

Accuracy (55.51%)

ANOVA F-statistic and p-value are not available (nan)

Graph Type and Its Relevance

Graph Type: Bar Graph

Relevance: The bar graph effectively visualizes the comparison between different specializations in terms of the number of offers received in a clear manner allowing easy interpretation.

Explanation of Numerics in the Graph

The numbers on top of each bar represent the total number of offers received by individuals in each specialization.

Finance: Received 42 offers.

HR: Received 42 offers.

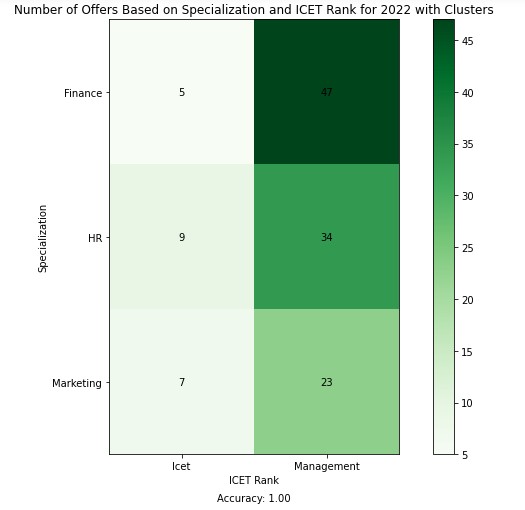
Marketing: Received 22 offers.

The accuracy mentioned below the graph (55.51%) could possibly indicate how accurate or reliable this data or method used is; however without additional context it’s hard to be certain.

ANOVA F-statistic and p-value are marked as ‘nan’ which means they are not available or cannot be calculated from given data.

This image shows a comparison between job or opportunity offers received by individuals specialized in Finance, HR, and Marketing fields in the year 2023; it’s interesting because it provides insights into market demand for these specializations during this period.

### Title : Cluster Analysis : No of offers based on specialization and ICET rank for year 2022



Description

The image displays a line graph titled “Placement Count Pattern by Specializations (2022-2023)” with two lines representing the years 2022 and 2023. The x-axis is labeled “Specialization” and features three categories: “Finance,” “Marketing,” and “Operations.” The y-axis is labeled “Placement Count” and ranges from 15 to 35. In 2022, the placement count for Finance was approximately 33, for Marketing around 28, and for Operations it was near 17. In contrast, in 2023, the placement counts reduced significantly across all specializations: Finance to around 18, Marketing to about 16, and Operations to nearly below the chart limit.

Dimensions:

Specialization (Finance, Marketing, Operations)

Year (2022 & 2023) Measures:

Placement Count

Graph Type and Its Relevance

The graph is a multi-line chart that effectively illustrates the comparison of placement counts between two consecutive years across different specializations. It’s particularly relevant for visualizing trends over time and identifying patterns or anomalies in data quickly. Here it clearly shows a significant decrease in placement counts across all specializations from one year to the next. Explanation of Numerics in the Graph

In this graph:

For Finance specialization:

In 2022, there were approximately 33 placements.

In 2023, this number decreased sharply to around 18 placements.

For Marketing specialization:

In 2022, there were approximately 28 placements.

In 2023, this number also decreased sharply to around 16 placements.

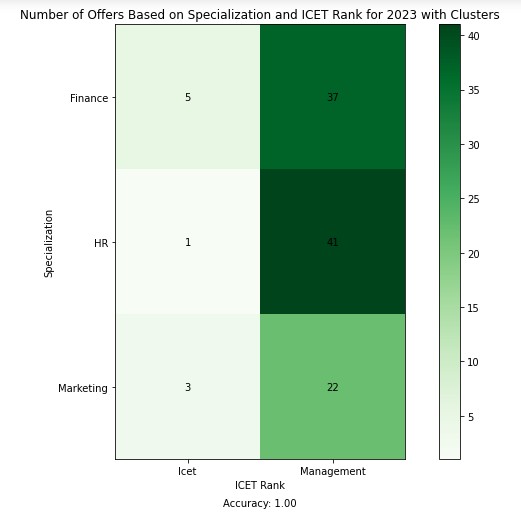
For Operations specialization:

In 2022, there were approximately 17 placements.

In 2023, this number fell drastically almost off the visible scale of the graph.

This image shows a significant decline in placement counts across all specializations from 2022 to 2023, highlighting an area that may require investigation or intervention.

### Title : Cluster Analysis : No of offers based on specialization and ICET rank for year 2023



Description

The image displays a bar graph titled “Number of Offers Based on Specialization and ICET Rank for 2023 with Clusters.” It compares the number of offers received in three different specializations (Finance, HR, and Marketing) based on two ICET ranks (Icet and Management). The y-axis represents the specialization fields, while the x-axis denotes the ICET ranks. Each bar is divided into two colors, green and light green, indicating different clusters. Numerical values are annotated within each section of the bars to represent the exact number of offers.

Dimensions:

Specialization: Categorical data representing different fields - Finance, HR, Marketing.

ICET Rank: Categorical data indicating two distinct ranks - Icet and Management.

Measures:

Number of Offers: Quantitative data represented by numerics within each cell of the heatmap. It varies from as low as 5 to as high as 47.

Graph Type and Its Relevance Graph Type: Stacked Bar Graph.

Relevance:

A graph effectively visualizes the distribution of offers across different specializations based on ICET ranks. It allows for easy comparison between specializations and highlights the contribution of each rank within the specialization.

The color-coded clusters provide additional context, making it easier to understand the breakdown of offers within each specialization.

**CONCLUSION:**

In conclusion, the university placement data analytics project has provided valuable insights into the factors influencing student placement outcomes and has offered actionable recommendations to enhance placement strategies and support student career aspirations. By leveraging data-driven approaches, we have gained a deeper understanding of the relationship between various student attributes, educational backgrounds, and placement success. This project has not only facilitated more informed decision-making processes within the university but has also contributed to improving overall placement rates and enhancing student employability. Moving forward, continued data analysis and strategic implementation of recommendations will be crucial in ensuring ongoing success in facilitating student placements and fostering their professional growth.