

Placement Analysis and Industrial Recruitment Trends in BITS, Pilani.



Final Report

Group - 6

Daggupati Sai Aravind Kasyap	2022H1540832P
Muttevi Sai Yasasvi	2022H1540831P
Chamarthi Sai Saran	2022H1540808P
Sindhe Pandurang	2022H1540810P
Sai Praveen Panuganti	2022H1540824P

Table of Contents

1. Problem Statement	3
2. Evolution	3
a. Statistics Project to Data Visualization Project	3
b. Data Collection	3
c. Data Sources	4
3. Architecture	5
4. Is it a Project or a Product	5
5. ETL Phase	6
a. Data Extraction	6
b. Data Transformation	7
6. Data Modelling	8
a. Data Analysis Expressions	9
7. Data Visualization	10
a. Visualization Reports and Insights	10
8. Business Perspective	15
a. Model and Value Proposition	15
b. Implementation	16
9. Future Scope	18

Problem Statement

This is an extension of the statistics project “Factors affecting final placements offers”. In this version of extension, data from different departments such as MBA, Pharmacy, MTech etc. were taken into consideration. As the aesthetics change with the change in departments, we gathered the raw data from all the sources such as google forms and the placement chronicles of the institute and performs ETL operations using the Power BI and visualizes it in a most insightful manner.

Evolution

Statistics project to Data Visualization project:

The Statistics project primarily focused on quantitative data to predict CTC based on various factors. However, for this project, our focus is on gathering significant insights related to the recruitment pattern from a macro perspective. Our emphasis is on industry perspectives and recruitment trends such as:

1. What is the role offered, are they dependent on the branch specialization or technical competencies of students or certifications.
2. Which technical competencies are associated with higher packages?
3. Which domains are recruiting the most people?
4. What does the industry expect in return for higher packages?

The expected outcomes of this project are based on data collected through a Google Form questionnaire. The survey includes questions related to quantitative data such as CGPA, CTC, and internships. However, the form primarily focuses on gathering qualitative information, including the sector of the company, role offered, technical competency details, etc. Furthermore, all branches of BITS have been included in the survey, providing more opportunities to identify differentiation in industry trends.


Data Collection:

As this project concerns placement analysis, the primary sources of data are obtained either directly from the placement cell or by distributing forms to final year students and gathering information.

We have floated the google form to the final year students of B Tech, M Tech and Pharmacy, whereas for MBA, data is being collected directly from the placement cell.

Data Sources:

1. Google Form survey
2. Placement cell

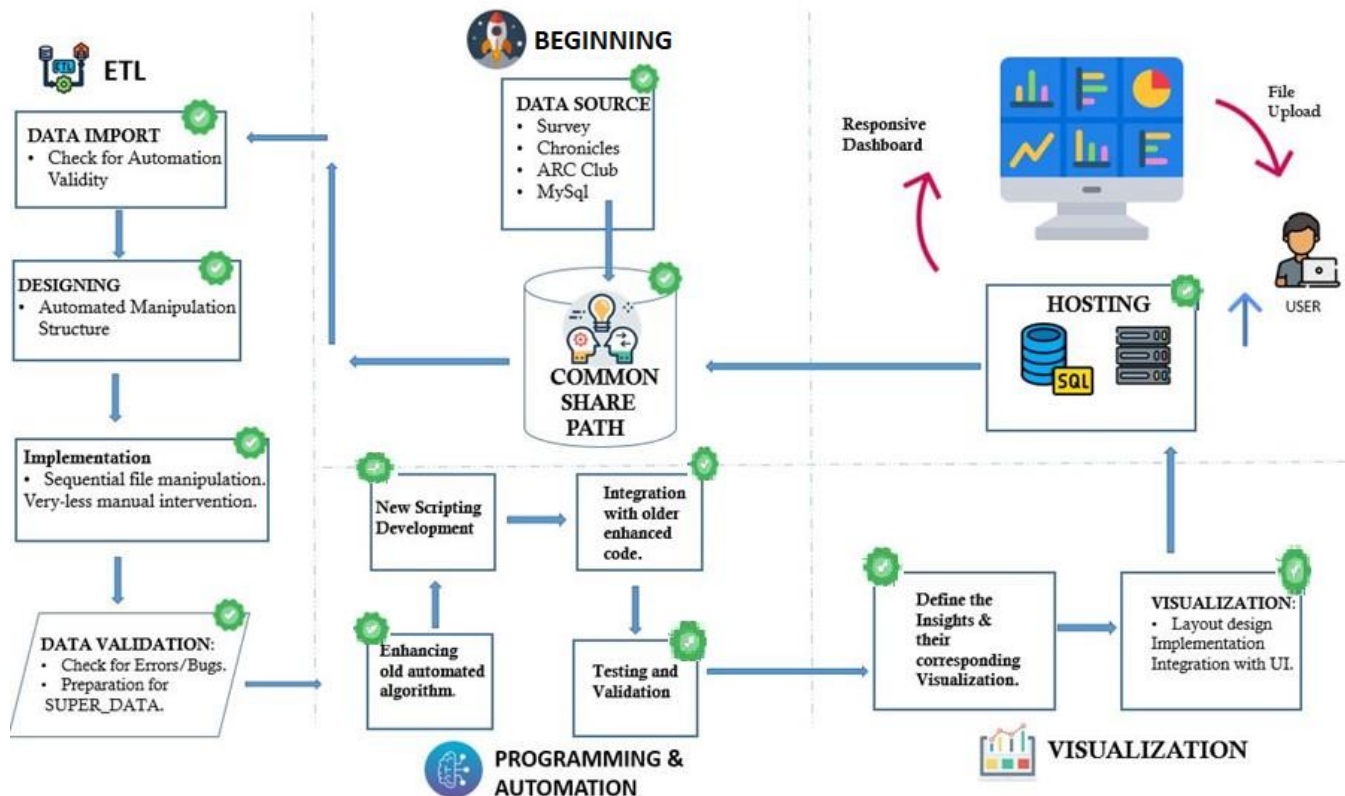


"Analyzing Placement trends & Scenarios across discrete Industry verticals."

- Many of us has have good grasp over the placement basic statistics like mean package or %le of the top and bottom. But it is not enough to get the accurate picture about where the market is heading , where the opportunity lies , to to know what are the key metrics(Variables) that have the highest impact on your placements.
- we are a team of five from management department(Business Analytics) working on a project to solve this problem i.e. **Information asymmetry**.
- After the Data Collection we will use discrete statistical models to find the accurate insights and after that we use Data Visualization tools to publish those insights on an open platform so that these can be accessible to all the **#Bitsians**.

The survey data and placement unit data are the primary sources of data required for this project. The insights gained from this data are valuable that they expand the project's scope beyond its original subject matter.

Architecture



Is it a Project or a Product?

Our product is designed to be versatile, capable of handling any placement-related data. Moreover, we have built a complete end-to-end framework that can convert an academic Data Visualization project into a functional product. By using automation and architectural expertise, we achieved our goal of transforming a Data Visualization project into a fully-fledged product.

Our architecture includes ETL, Modeling, and Visualization, with windows automation tools like Appium, all working together to create a complete end-to-end product or application.

ETL phase

➤ Data Extraction: Importing Files from folder

Combining 3 different files from the folder

1. Google form
2. Placement unit
3. B Tech placement unit

Problem: Different header names in all three files

Solution: Creating an Excel file (Header conversion) with source and target header names importing that header conversion excel file with source and target header names in transform sample file

- Demote the current headers
- Transpose the data
- Merge queries based on source names from transformation sample File and header conversion file
- Expand the data in merged queries by selecting only the target
- Add a conditional column to replace all the null values of header conversion table with sourcenames from transform sample file
- Remove source names column from transform sample file and move the merged column to starting column as to be used for header names
- Transpose back to get the combined data from all the sources with correct header names

Transform File

```
let
    Source = (Parameter1 as binary) => let
        Source = Csv.Document(Parameter1, [Delimiter=","], Columns=22, Encoding=1252, QuoteStyle=QuoteStyle.Csv),
        #Promoted Headers = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
        #Demoted Headers = Table.DemoteHeaders(#Promoted Headers),
        #Changed Type = Table.TransformColumnTypes(#Demoted Headers, {{"Column1", type text}, {"Column2", type text}, {"Column3", type text}, {"Column4", type text}, {"Column5", type text}, {"Column6", type text}, {"Column7", type text}, {"Column8", type text}, {"Column9", type text}, {"Column10", type text}, {"Column11", type text}, {"Column12", type text}, {"Column13", type text}, {"Column14", type text}, {"Column15", type text}, {"Column16", type text}, {"Column17", type text}, {"Column18", type text}, {"Column19", type text}, {"Column20", type text}, {"Column21", type text}, {"Column22", type text}}),
        #Transposed Table = Table.Transpose(#Changed Type),
        #Merged Queries = Table.NestedJoin(#Transposed Table, {"Column1"}, header_conversion_table, {"Source"}, "header_conversion_table", JoinKind.LeftOuter),
        #Expanded header_conversion_table = Table.ExpandTableColumn(#Merged Queries, "header_conversion_table", {"Traget"}, {"header_conversion_table.Traget"}),
        #Reordered Columns = Table.ReorderColumns(#Expanded header_conversion_table, {"header_conversion_table.Traget", "Column1", "Column2", "Column3", "Column4", "Column5", "Column6", "Column7", "Column8", "Column9", "Column10", "Column11", "Column12", "Column13", "Column14", "Column15", "Column16", "Column17", "Column18", "Column19", "Column20", "Column21", "Column22"}),
        #Added Conditional Column = Table.AddColumn(#Reordered Columns, "final_columns", each if [header_conversion_table.Traget] = null then [Column1] else [header_conversion_table.Traget]),
        #Reordered Columns1 = Table.ReorderColumns(#Added Conditional Column, {"final_columns", "header_conversion_table.Traget", "Column1", "Column2", "Column3", "Column4", "Column5", "Column6", "Column7", "Column8", "Column9", "Column10", "Column11", "Column12", "Column13", "Column14", "Column15", "Column16", "Column17", "Column18", "Column19", "Column20", "Column21", "Column22"}),
        #Removed Columns = Table.RemoveColumns(#Reordered Columns1, {"header_conversion_table.Traget", "Column1"}),
        #Transposed Table1 = Table.Transpose(#Removed Columns),
        #Promoted Headers1 = Table.PromoteHeaders(#Transposed Table1, [PromoteAllScalars=true])
    in
        #Promoted Headers1
in
    Source
```

➤ **Data Transformation:**

Problem 1: Source file name, feedback column, disciplined scale is present in data which is to be removed.

Transformation used: “Remove columns” function from Home Tab

Solution: Remove source.name column, feedback column, disciplined scale using transformtab.

Problem 2: Values given as “-” for the unplaced students from the data source, those are unplaced

Transformation used: “Filter function” option using Transform Tab.

Solution: Filter out the rows which are filled “-” in CTC field

Problem 3: Communication Scale, Discipline, Competitions won, No.of Domain Tech, have null values due to non-mandatory fields during survey form

Transformation used: “Replace Values” using Transform Tab

Solution: Replaced the null values with a minimum expected number for the fields as, Communication Scale – 5, Discipline - 0, Competitions won - 0, No.of Domain Tech – 0 respectively.

Problem 4: Proficient Coding languages, Internship Domain columns are multi valued attributes

Transformation used: “Split Column” using Transform Tab

Solution: Filter out the rows which are filled “-” in CTC field

Problem 5: Presence of Trailing/ Leading spaces in columns values

Transformation used: “Format” using Add Transform Tab

Solution: Using Trim option remove the white spaces of the columns

Problem 6: Exceeding 2-3 internship domains

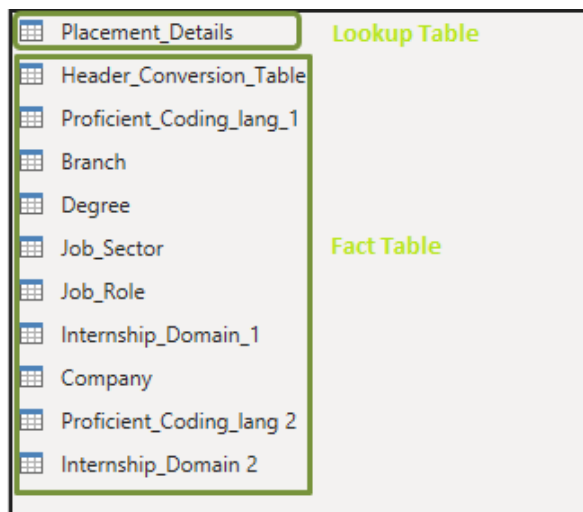
Transformation used: “Remove Columns” using Home Tab

Solution: Only the top two internship domains were analyzed, and any split columns with more than 2 were removed

Problem 7: Difficulty in modelling and visualizing data due to a single table format.

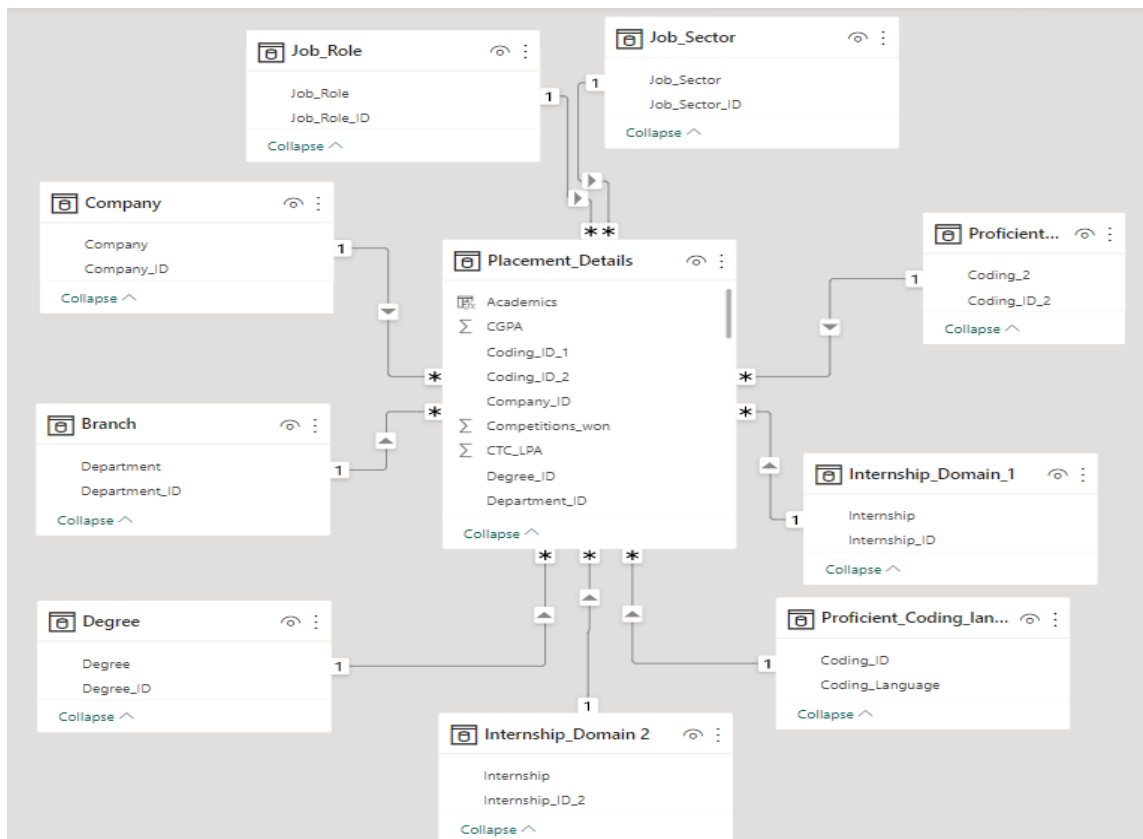
Transformation used: Right click on the column and select “Add a new Query”

Solution: Split the single table into lookup table and fact tables with respective IDs



Data Modelling

Establish a *Model View* as below Using the lookup and fact tables.



➤ Data Analysis Expressions

- Below measures are created using DAX

Calculated Columns

- Academics: Categorized academic performances as Excellent, Good and Average based on CGPA**

```
Academics = IF(Placement_Details[CGPA]>8.5,"Excellent",IF(Placement_Details[CGPA]>7.5,"Good",
"Average"))
```

- Skills: Skill rating was used to categorize skills as High, Decent, and Average.**

```
Skills = IF(Placement_Details[Skill_Rating]>8,"Highly Skilled",IF(Placement_Details[Skill_Rating]
>=7,"Decently Skilled","Average Skilled"))
```

- Total Internships: Classified Internships into Single, Multiple, None**

```
Total Internships = IF(Placement_Details[Internships]>1,"Multiple Internships",IF(Placement_Details
[Internships]=1,"Single Internship","No Internship"))
```

Other Measures

Below are the few measures used for visualization

- Average Certifications**

```
Average Certifications = AVERAGE(Placement_Details[No_of_Certs])
```

- Average CGPA**

```
Average CGPA = AVERAGE(Placement_Details[CGPA])
```

- Average Coding Languages**

```
Average Coding Languages = AVERAGE(Placement_Details[No_of_Coding_Lang])
```

- Average CTC**

```
Average CTC = ROUND(AVERAGE(Placement_Details[CTC_LPA]),2)
```

- Average Internships**

```
Average Internships = AVERAGE(Placement_Details[Internships])
```

- Average Non-Tech Certifications**

```
Average Non Tech Certifications = AVERAGE(Placement_Details[No_of_NonTechCerts])
```

- Average Skill Rating**

```
Average Skill Rating = AVERAGE(Placement_Details[Skill_Rating])
```

- Average Technical Certifications**

```
Average Technical Certifications = AVERAGE(Placement_Details[No_of_TechCerts])
```

- Highest CTC**

```
Highest CTC = MAX(Placement_Details[CTC_LPA])
```

- Internship Total**

```
Internship Total = SUM(Placement_Details[Internships])
```

- Least CTC**

Least CTC = MIN(Placement_Details[CTC_LPA])

- **Median CTC**

```
Median CTC = MEDIAN(Placement_Details[CTC_LPA])
```

- **Number Of Offers**

```
Number Of Offers = DISTINCTCOUNT(Placement_Details[Student ID])
```

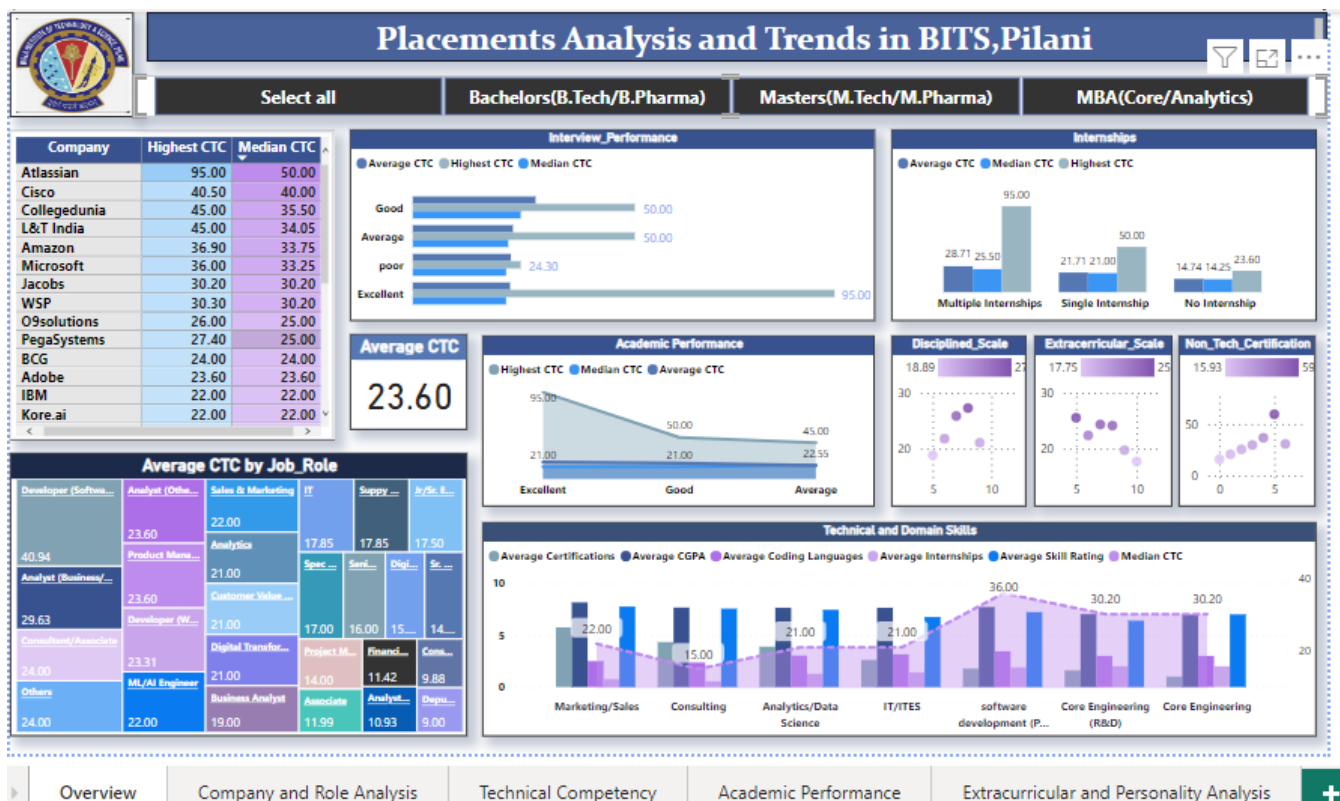
- **Total Companies**

Total Companies = `DISTINCTCOUNT(Placement_Details[Company_ID])`

Data Visualization

Visualization Reports and its Insights

- **Placement Analysis Overview**



Placements analysis Insights

- General overview of companies, roles and package range:
 - Companies that specialize in core technologies such as software product development, cloud computing, networking software, and data science, like Microsoft, Amazon, Altasian, and Cisco, offer high median packages ranging from mid 30s to even 50 lakhs.
 - After tech, the next in line are core engineering companies like L&T and Jacobs that offer

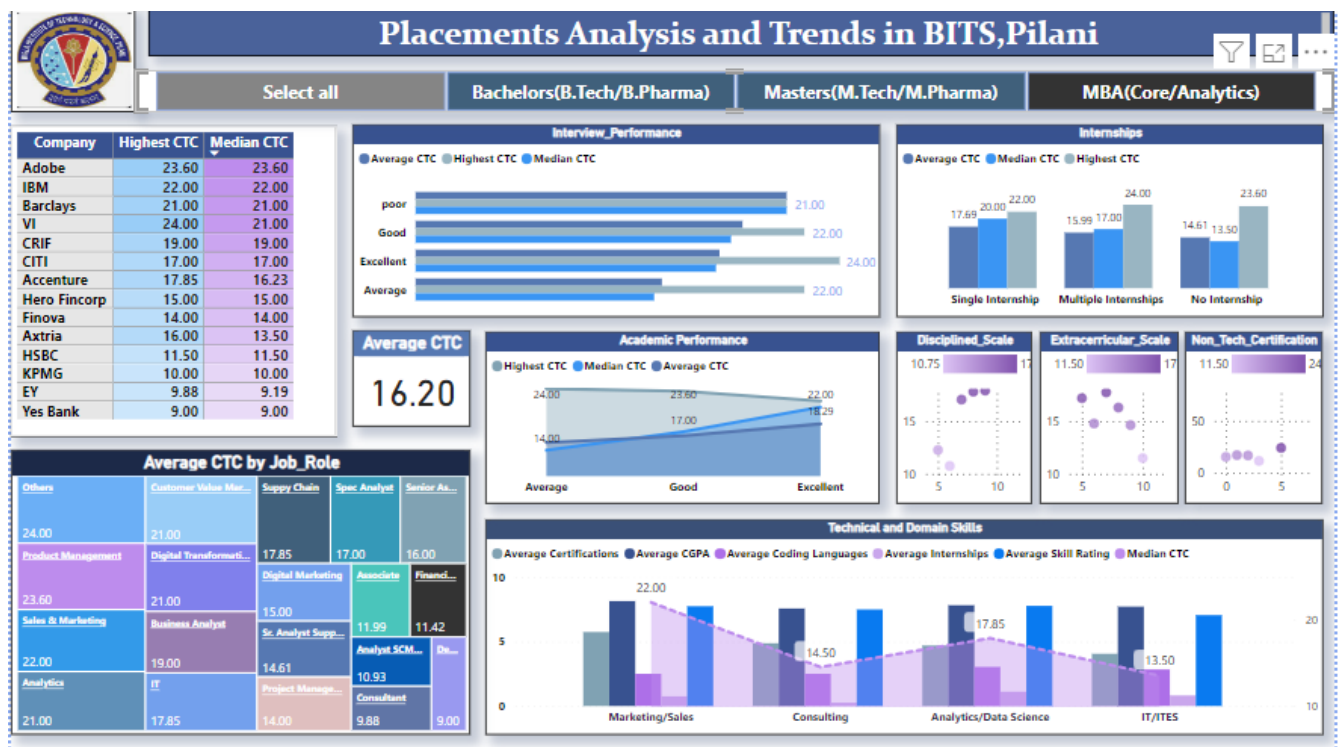
software developer or data scientist roles.

- Consulting companies like BCG and WSP are next in line, offering median packages in the range of 25-30 LPAs, primarily for associate, junior consultant, or analyst roles.
- Start-ups like 09solutions and Flipkart also offer similar packages with offers ranging from developer/analyst to a data scientist, with typical median package ranges from 20 LPAs to 30 LPAs.

➤ Effect of Interview Performance on Placements (B.tech):

- The mean CTC offered had a direct positive linear relation with the interview performance, i.e. better the performance better the Median CTC is.

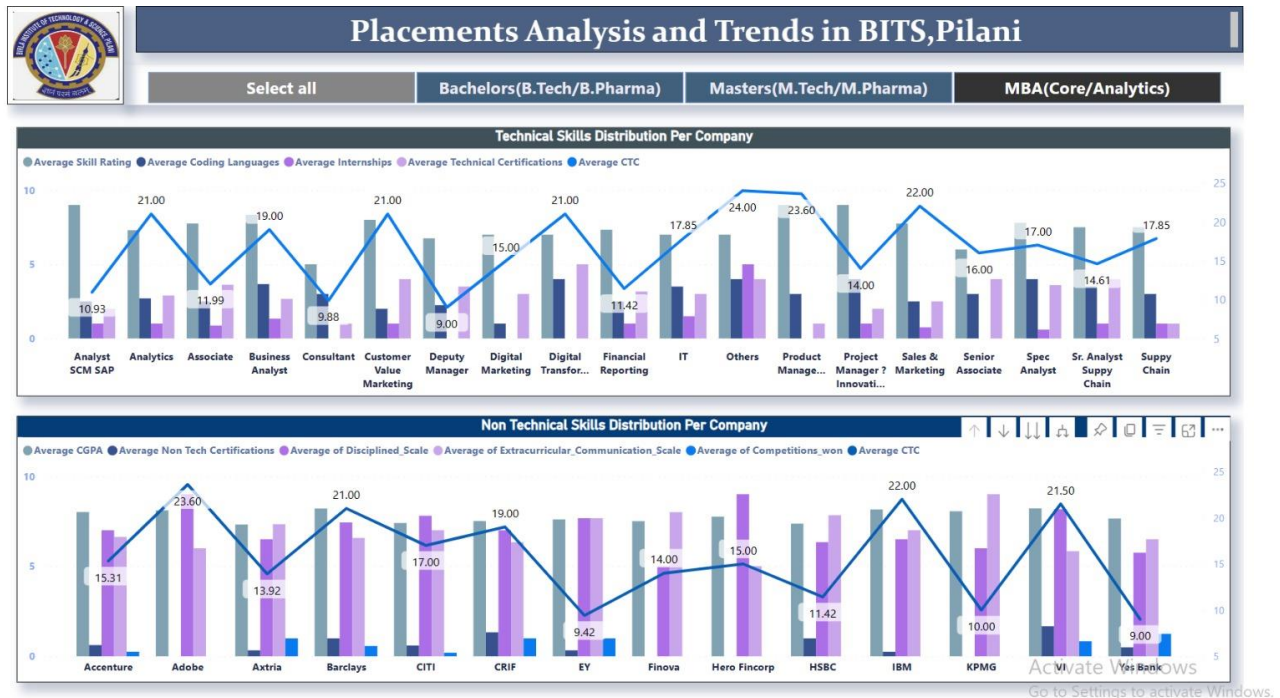
• **Placement Analysis Overview – MBA**



➤ Effect of Interview Performance on Placements (MBA):

- For MBA, it's the other way around when compared to B.tech, we believe it could be because MBAs have high degree of variables compare to B.tech like past experience, more no of certifications, etc. So, especially for MBAs the overall skill level and background has higher weightage of CTC as compare to B.tech

- Company vs Role Analysis



➤ CTC offered w.r.t to roles and domains (b.tech):

- In terms of roles, core technological roles such as software developers fetch the highest median package, which is around 42 LPA. However, we believe that this figure is highly inflated by the highest packages offered in software development.
- Data science and analyst roles come in second, with a median package ranging from 24 to 30 LPA. As the market experiences a data science boom, we believe these packages reflect this trend.
- Consulting, product management, and marketing roles are next in line, with a median package ranging from 21 to 26 LPA. Consulting roles offered by blue-chip companies like BCG, McKinsey, and WSP, along with product management roles offered by start-ups and marketing roles offered by FMCG companies, fall into this category.
- Core engineering and entry-level management roles are at the bottom of the category, with packages ranging from 10 to 19 LPA. Though these roles are offered by established core companies, they offer high job stability despite lower compensation packages.

➤ CTC offered w.r.t to roles and domains (MBA):

- When it comes to MBA, the sales and marketing roles take the lead, with a median package ranging from 15 to 22 LPA. The primary roles offered are sales specialists, customer value marketing, and digital marketing.
- The analyst roles are second after sales, with a median package ranging from 15 to 21 LPA and

roles ranging from senior business analyst to finance analyst. Nearly 50% of the 2021-23 MBA batch got placed in the analytics domain, and the front-runner companies in this sector are financial behemoths like Barclays, Citi, HSBC, Yes bank, and other companies like VI, IBM, and Atria.

- Consulting fetched the lowest median CTC compared to the former two, ranging from 10 LPA to 15 LPA. EY, KPMG, and Accenture led the hiring in the consulting domain, offering roles such as associate or junior consultant, along with some analyst roles.
- Product development companies are giving the highest weightage to skill level. As product development roles are limited, companies are looking for the best of the best.

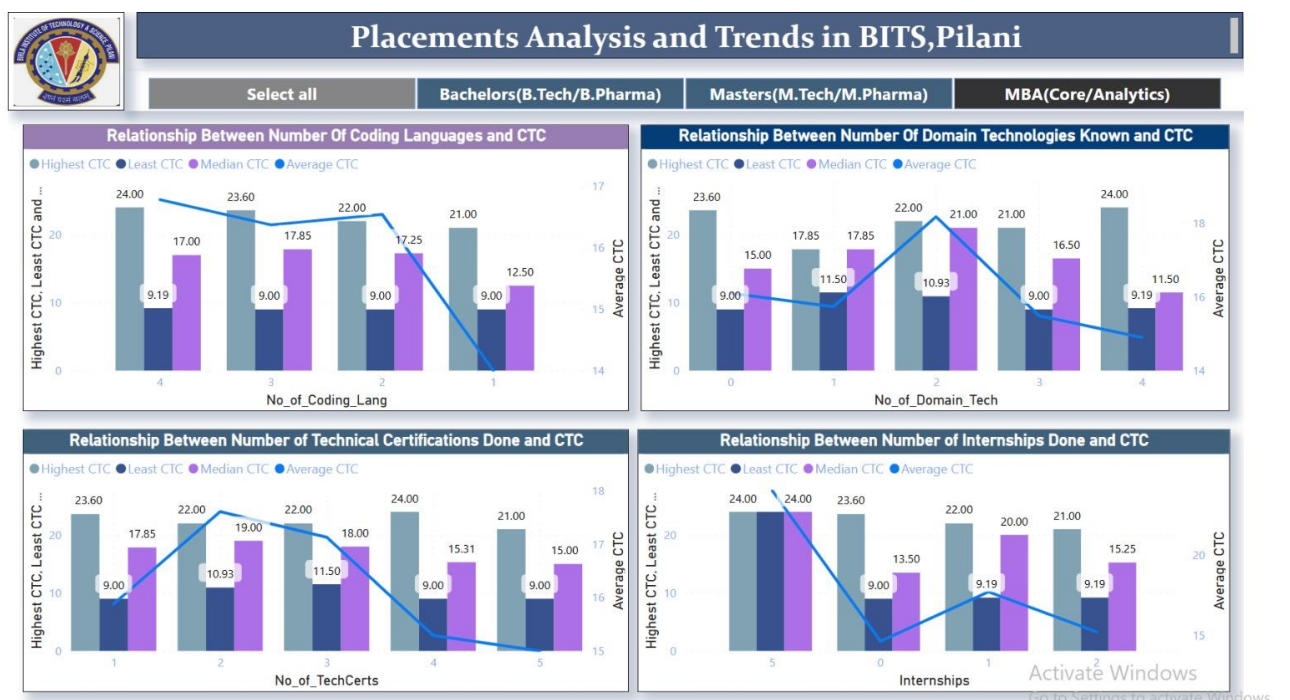
➤ Most desired skills by companies (B.tech):

- Companies that offer software and data science roles looking for people with proficiency in multiple coding languages and multiple technical certifications.
- Consulting companies focus on hiring graduates with more non-technical certifications and high communication skills.
- Companies hiring for sales and marketing roles focus more on communication skills compared to others, as marketing with communication is like playing FIFA without a joystick.

➤ Most desired skills by companies (MBA):

- Specifically for MBA, companies offering core analytics and data science roles, such as CIT, Barclays, and HSBC, look for people with overall technical and non-technical personalities, including a high number of certifications and communication skills.

• **Technical Competency**



➤ Effect of Multiple Internships (B.tech & MBA):

- The common myth is that those who have multiple internships are more likely to get good packages, but our finding tells that there is hardly any correlation between those two. In fact, we found the opposite.
- In B.tech grads with multiple internships, the median package was 10% lesser than those with a single internship.
- For MBA, it's even worse. The multiple interned MBAs got 15% less package, i.e., 17 LPA median, compared to the single intern who got 20 LPA.

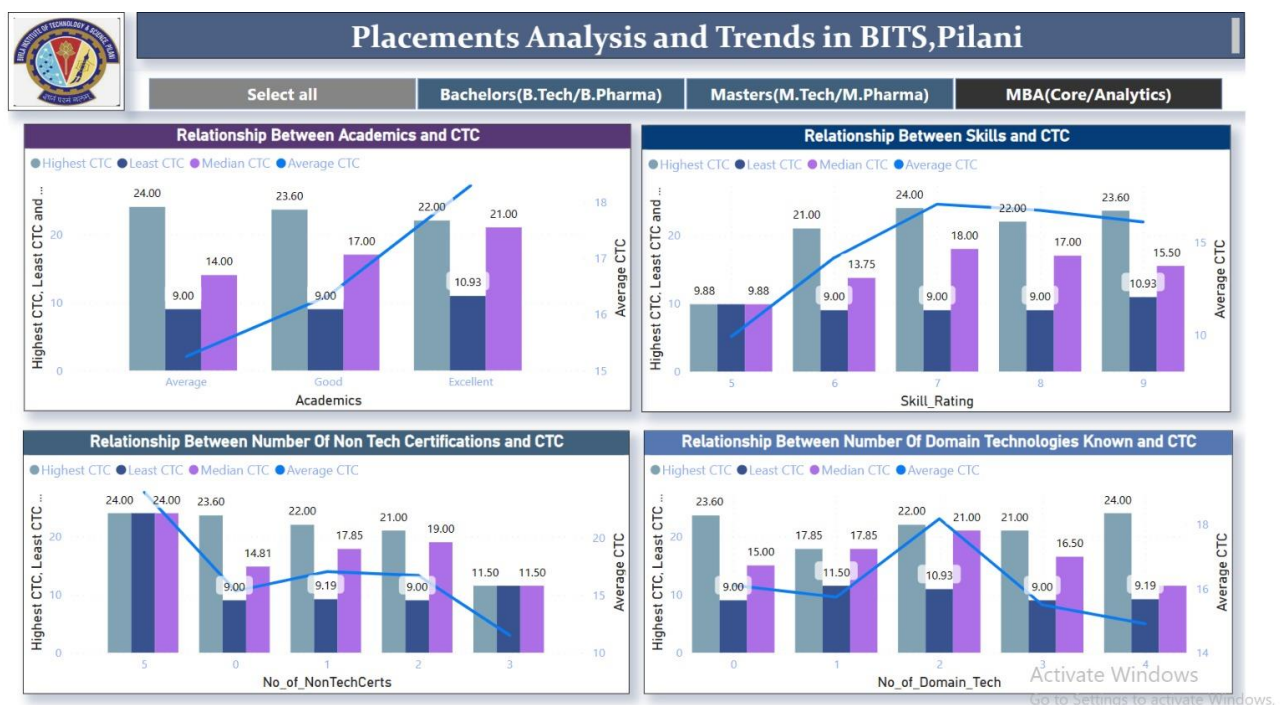
➤ Effect of Technical/Domain Skills (B.tech & MBA):

- Certifications: Companies offering roles in streams such as marketing/sales, consulting, and analytics/data science place greater importance on the number of relevant certifications you have compared to those in core technological or engineering fields.
- Number of coding languages: Proficiency in multiple coding languages has significantly increased the median salary, especially for roles in software and data science domains. Therefore, it is recommended that you improve your skills in multiple coding languages, including databases.
- Average internships: Although having multiple internships may reduce the median salary by 10-15%, it is positively correlated with the median salary for specific instances such as data science and analytics.

➤ Effect of Technical Competency on CTC (B.tech & MBA):

- Our analysis shows that having multiple technical certifications, fluency in multiple coding languages, and domain technologies can significantly boost one's CTC.

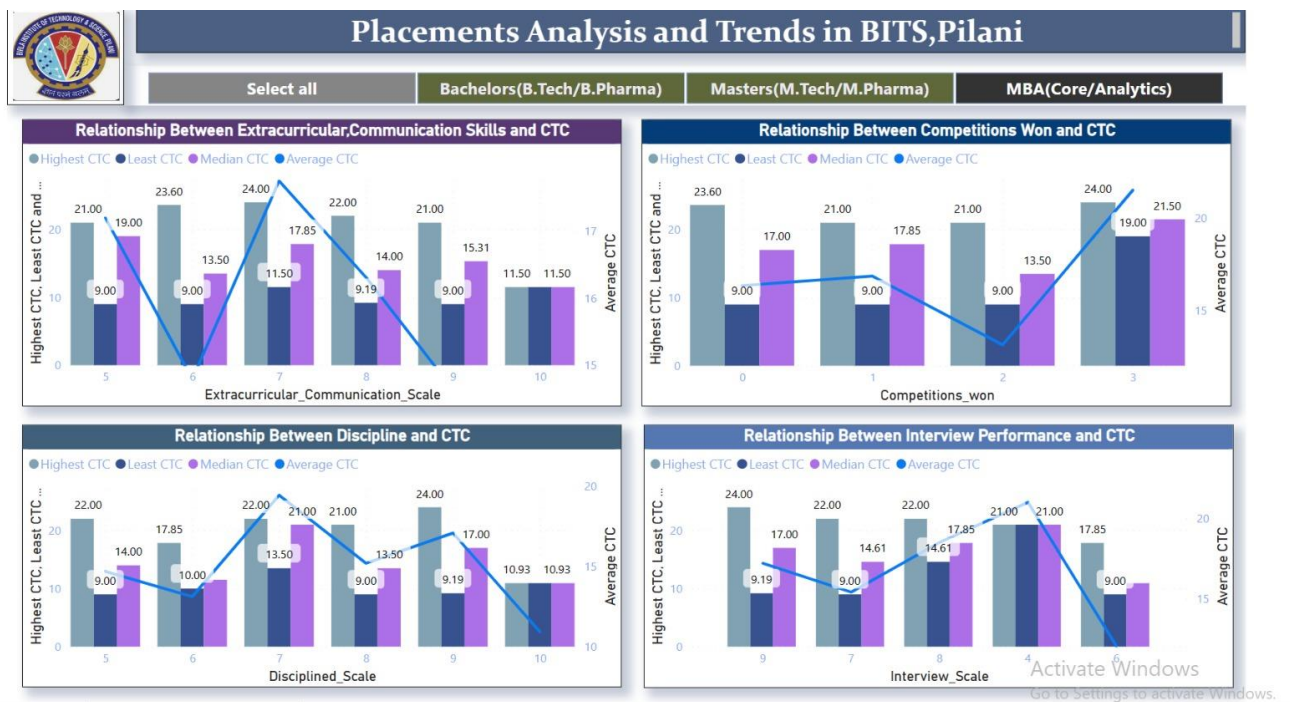
• **Academic Performance**



➤ Significance of Academic Performance (B.Tech & MBA):

- If we visualize the whole MBA and B.Tech grad's academic performance vs. median CTC they receive, one thing is clear. The CTC gain over CGPA increment is not significant. That is, there is hardly any difference between median CTC received by those with an average vs. excellent CGPA.
- Our finding suggests that studying just enough to stay above average and make efficient use of the time for other productive tasks such as developing your technical proficiency or sharpening your business skills.

• **Extracurricular and Personality Analysis**



➤ Extra-curricular and personality analysis (B.tech & MBA):

- Individuals who have won multiple business competitions tend to receive significantly higher median packages compared to those who have won less or none.
- However, the opposite is true for individuals who possess extracurricular skills such as communication. Therefore, it is recommended to balance both technical and communication skills to increase overall employability and potential salary.

Business Perspective

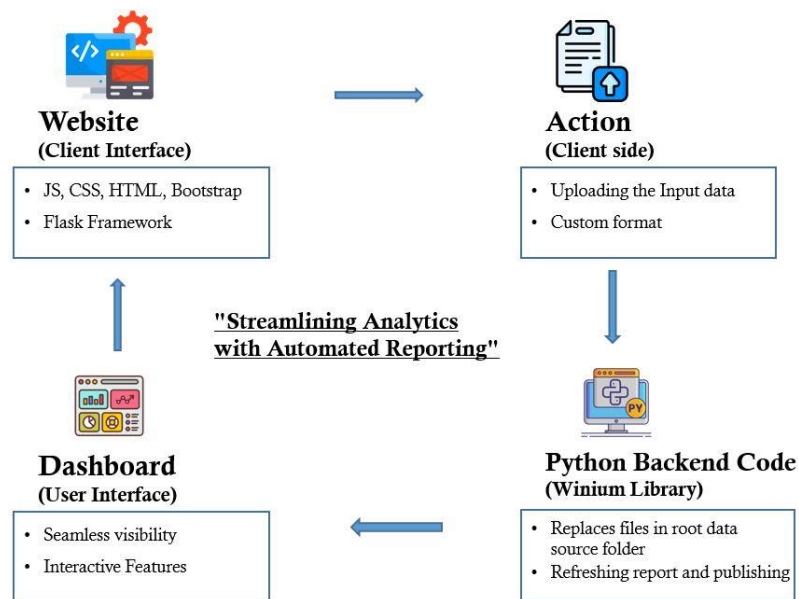
➤ **Model and Value Proposition**

An end-to-end application that offers interactive data visualization and insights based on placement data. The application's value proposition is its ability to deliver engaging and visually immersive analytics using the inputted placement data, It provides users with a unique and

experience through an interactive platform to analyze trends in hiring, as well as the factors that impact role offering and placement packages.

The target audience for the application will be educational institutions seeking data-driven insights to inform their decision-making towards their placement processes.

➤ Implementation



About the Product: Interactive application where the user has an option to input placement data in a required format to the application for getting the insights and analysis of the data resulting in a visually interactive dashboard on UI

Application URL: <https://aravindkasyap.github.io/PO/>

Technologies Used: Power BI, Winium for automation, Flask (python web framework), Git hub for hosting the application

Application frontend overview

InsightX

Home App Features Upload Dashboard

Focus on what matters

You don't need to work hard to earn an empire; there is an army of slaves to do it for you.

Upload

Dashboard

We try to create a viable product that can be used across similar applications, platforms and requirements

Placement data that can be analyzed providing insightful information to improve and focus on what matters i.e. Academics or Technical or Non technical or Extra curricular or a combination of any.

using every functionality to the fullest to induce the best results and insights dynamically with very minimal manual touch points maintaining the data integrity across the complete process.

- ✓ Single dashboard contains different perspectives of placements.
- ✓ Insightful representations of Academic, Technical, Non technical versions of the same data
- ✓ Easy application of filters across all the visualizations

App Features

Need a quick and easy way to check placement insights and industry trends? Have a look at seamless and automated architecture over

Template Uniformity

Predefined template allows unified input pattern allowing no discrepancies at any stage of processing.

Automated

application backend is completely automated including the preprocessing and post response validations.

Secure

File nullification in post response options and dashboard's secure embedment

Interactive dashboard

Application of filters and sliders directly on the visualizations without latency

Swift UI

Most seamless and smooth navigation between sections and in page reloads

responsive

Uniform and presentable UI across all platforms of devices

Upload input files

upload input files as per the template given below

Multiple files can be uploaded using the input template format and click on submit to process the input. data and visualize the input. Dashboard will be visible in dashboard section of the page

Download input template

Choose Files

No file chosen

Submit

Dashboard

click on the visualizations to apply any filters

Overview

Company and Role Analysis

Technical Competency

Academic Performance

Educational and Tenability Analysis

Designed by Kaseya

© Copyright Group & All Rights Reserved

Future Scope

Integration of the statistics project "Determining the factors affecting final placement offers (CTC) using multiple linear regression" into this application. The integration enables the end-user to not only observe the current placement data trends but also forecast the underlying factors responsible for favorable placement offers.