



PROJECT REPORT

DATA ANALYTICS  
  
Career Analytics with Python

|  |  |  |  |
| --- | --- | --- | --- |
| **Created By:** | Krish Desai | **Approved By:** | Harshada Topale |
| **Created On:** | 23-07-2024 | **Approved On:** | DD-MMM-YYYY |

Page left blank intentionally

**INDEX**

**[1](#_Toc143445375)****[PROJECT DETAILS](#_Toc143445375)** [5](#_Toc143445375)

**[2](#_Toc143445376)****[SUMMARY](#_Toc143445376)** [5](#_Toc143445376)

**[3](#_Toc143445377)****[INTRODUCTION](#_Toc143445377)** [5](#_Toc143445377)

[3.1 Background 5](#_Toc143445378)

[3.2 Stakeholders 5](#_Toc143445379)

[3.3 Objectives 6](#_Toc143445380)

**[4](#_Toc143445381)****[METHODOLOGY](#_Toc143445381)** [6](#_Toc143445381)

[4.1 Considerations & Assumption 6](#_Toc143445382)

[4.2 Approach 6](#_Toc143445383)

[4.3 Activities 7](#_Toc143445384)

**[5](#_Toc143445385)****[TARGETTED V/S ACHIEVED OUTPUT](#_Toc143445385)** [7](#_Toc143445385)

**[6](#_Toc143445386)****[CONCLUSION](#_Toc143445386)** [8](#_Toc143445386)

**[7](#_Toc143445387)****[APPENDICES](#_Toc143445387)** [9](#_Toc143445387)

[7.1 Appendix A – Title 9](#_Toc143445388)

**General Instructions for using the Live Project Report Template**

* This template and the subsequent document created using this template is a confidential document and is the intellectual property of Cloud Counselage Pvt. Ltd. Circulating it outside of the organisation without the consent of Cloud Counselage Pvt. Ltd. is the breach of company policies and will lead to legal actions
* This template is a guideline document to communicate the implementation of design ideas and the results of the work to the stakeholders.
* The **text between inequality (< >) is to be replaced** by relevant text
* Please **remove the yellow highlight on the Text** between the inequality (< >). This is done to help you notice the text to be changed/replaced
* The text in *italics* highlighted in grey is just for reference and should be removed after adding the relevant text

# **PROJECT DETAILS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Project Name** | Career Analytics with Python | | |
| **Project Sponsor** | Tushar Topale | | |
| **Project Manager** | Harshada Topale | | |
| **Start Date** | 05-06-2024 | **Completion Date** | 25-07-2024 |

# **SUMMARY**

The project aimed to give a clear understanding of the demographics of the students who will be applying for jobs. It gives insights into the relation between academic and non-academic factors of an individual with their job expectations.

This projects helps in understanding the self-awareness of the students and help recruiters, counsellors and educational instituions make informed discisions based on the valuable insights found in this project.

The project ultimately aims to rove facts and validate certain assumptions that we have about students and their career aspirations.

# **INTRODUCTION**

## Background

Millions of students apply for internships/jobs every year. Resume play an important role in playing the first impression. The recruiters spend a max 2-3 minutes reviewing a resume after it landed in their mailbox or job board, ATS application. Surprising more than 70% of resumes get rejected in the initial screening.

This project focuses on understanding the backgrounds of students in regards for their professional careers. Taking into account the financial background and academic achievements of each student and their relationship(if any) with their job expectations.

## Stakeholders

Students : Proper guidance and understanding potential.

Recruiters : Get an great idea about the skills and merit of new and future applicants

Educational institutions : Helps them understand the student demographic

Counsellors : Gives the insights into the profiles of students about to start their professional career

## Objectives

The objective is to find patterns, trends and establish a relation(if any) between the resumes i.e their academic achievements, financial backgrounds and their future goals and expectations.

# **METHODOLOGY**

These conventions are all about the positions of line breaks, how many characters should go on a line, and everything in between.

## Considerations & Assumption

1. If the collection of data is not systematic, feature selection becomes complex and pre-processing takes a lot of time. A lot of redundancy in data affects the analysis.
2. Since python is being used for analysis, scaling up may be difficult. Python make not work as efficiently with a very large dataset.
3. The metrics that are used are similar and can be compared. Differents metrics used in different regions to score a students academic achievements may lead to incorrect analysis and unreliable results.

## Approach

The following approach was implemented to solve the problem statement :

**Data pre-processing:**

The dataset had been cleaned for missing values, outliers and unnecessary data. As the data collection was not perfect some feature engineering was performed.

**Data analysis :**

Python was used for performing EDA and analytical work(statistical/qualitaive)

**Visualization :**

Python was used for visualization. The results of the analysis and EDA were presented in graphs and diagrams for better understanding.

## Activities

## Requirement gathering and planning :

In this phase we collect the data and asses the quality of the data. We understand and study the problem statement to lay down our goals and expected results. We make a systematic schedule and breakdown the project into achievable milestones with probable deadlines.

**Preparation :**

In this phase we perform data cleaning and pre-processing of the data to make it ready for analytical and visualization work & activities.After studying the problem statement we get rid of irrelevant data and values.We also derive some required data from the collected data. Includes data transformation.

**Execution :**

In this phase we solve the problem statement using Python and its various libraries. We achieve the results and solutions. This includes statistical and qualitative analysis. Once we have our results and findings we create visualizations for presentation purposes.

**Documentation :**

In this phase we make clear and precise documentation for our project as deliverables to be submitted. This includes SRS document, RAID logs, WBS(Work Breakdown Structure), lessons learnt. WE also document the technical aspects like code files. We generate a report and presentation for the stakeholders.

**TARGETTED V/S ACHIEVED OUTPUT**

**TARGETED OUTPUT :**

The primary aim of this project was to a comprehensive understanding of the demographics of the students who will be applying for jobs after graduation.

The project also aimed to give an insight on how academic and non-academic factors affect(or not) the job expectations.

The project aimed for the following :

1. Improve the self-awareness of students regarding their career prospects.
2. Help recruiters, counselors, and educational institutions to make informed decisions based on the insight s and results.
3. Validate the assumptions and establish patterns between students’ academic achievements, financial backgrounds, extra-curricular activities and future career goals.

**ACHIEVED OUTPUT :**

**The project successfully achieved its targeted objectives by:**

1. The project provided valuable insights into the relationship between students' academic and non-academic backgrounds and their job expectations, helping stakeholders in making informed decisions.
2. The use of Python for data analysis and visualization enabled the identification of significant patterns and trends. The results were effectively communicated through various graphical representations.

* The students got a better understanding of their potential and received guidance.
* The recruiters a deeper insight into the skills and merits of potential candidates, improving the selection process.
* The project helped educational institutes make better educational programs for students to better prepare them for the future careers.
* Counselors could give better career advice.

# **CONCLUSION**

As the name of the project suggests, this project gives an insight into the career prospects of freshly graduate students. It helps the stakeholders understand the demographic of students and potential new job applicants. The project gives valuable insights in the academic achievements/performance of students and its relational with their non-academic attributes like financial background, region and extra curricular activities. This project shows and highlights the prowess of Data Analytics in the educational field.

The project can be further improved by collecting a much larger data-set with even more features and attributes. With a much larger data-set the descriptive analytics can be taken to next level with more details,credibility and quality to the outcomes and insights.

# **APPENDICES**

## Appendix A – Title

Libraries used - pandas, matplotlib and seaborn

Dataset - provided by cloud counselage