

# **Acknowledgement**

We would like to express our special thanks of gratitude to our Teacher Mr. Rakesh Jha, who gave us the golden opportunity to do this wonderful project on Employment Analysis.

Who also helped us in completing our project. We came to know about so many new things.

We are really thankful to them. Secondly, we would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame.

**Aadarsh Prem,  
Ajit Ekka, and  
Adithya Anil  
BCA 2020**

# Content

- Introduction of the Project
- Objective of the Project
- SRS :- Hardware and Software requirement  
Introduction about the technology
- Database Link
- Code and Screenshot
- Software Testing
- Security
- Future Scope
- References

# Introduction of the Project

In this present project, we are working on a large amount of employee's dataset. This dataset is given in an excel file. We firstly corrected the formatting first so we divided the data in tuple (row and col) format. After that, we load or fetch data in a Data frame by using pandas library. We use pandas because a thousand or millions of data if you want to clean the messy data, analyze the data and manipulate the data according to your needs easily.

**Employee Database:** In this given project. In the data set of employees, there are different data types that are used in different columns of the table. Like employee ID data is in integer format, employee name and email in a character format, integer data type are used in a phone number column and also in Hiring date, Salary, Commission, Manager Id and department Id these are all integer data.

**Fetch Details of Employee:** the given dataset in an excel format and divided into different dataset as different sheets like: employee, Job grade, Countries, Job history, Jobs, department, location and Region. We fetch all the data in Python by using pandas. For fetching the Excel dataset into pandas we use import keyword, without import keyword we cannot use pandas or any library function. So we using "import pandas as pd"

**Python Pandas:** pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

# Objective of the Project

**Fetching the Data:** Pandas is extremely useful and comes in handy when we want to load data from various file formats like CSV, Excel etc.

**Data Cleaning:** When working with multiple data sources, there are many chances for data to be incorrect, duplicated, or mislabeled. If data is wrong, outcomes and algorithms are unreliable, even though they may look correct. Data cleaning is the process of changing or eliminating garbage, incorrect, duplicate, corrupted, or incomplete data in a dataset. The motive of data cleaning services is to construct uniform and standardized data sets that enable data analytical tools and business intelligence easy access and perceive accurate data for each problem.

**Increase searching ability:** Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

**Finding null or not null:** While making a Data Frame from a csv file, many blank columns are imported as null value into the Data Frame, which later creates problems while operating that data frame. Pandas `isnull()` and `notnull()` methods are used to check and manage NULL values in a data frame.

**Sorting operations:** Sorting is yet another pandas operation that is heavily used in data analysis projects. `sort_values()` method is used to perform sorting operation on a column or a list of multiple columns.

# **Software Requirement Specification (SRS)**

## **Introduction**

In this present project we are working on a large amount of employees dataset. This dataset is given in an excel file. We firstly corrected the formatting first so we divided the data in tuple (row and col) format. After that, we load or fetch data in a Data frame by using pandas library. We use pandas because a thousand or millions of data if you want to clean the messy data, analyze the data and manipulate the data according to your needs easily.

- **Purpose**

The purpose of our project is to establish and document the 'HR Database' of employment procedures such as Salary, Department\_ID, Location etc.

- **Scope**

The future scope for our project's extensive use of Big data in India ensure high employment, increases the compensation and helps individuals to connect with the advanced technologies. Our project can radically change the present business situation by capturing a large volume of data, expand business model, and energize the imaginative procedures and overall growth and development of a company.

- **Definition**

Our project aids companies and government organizations to collect data and identify patterns in the data. This helps to generate better insights for the organization about decision-making, and sometimes, automates the decision-making process itself. Our project can help to transform business and reduce the cost and time to make quick decisions. The data analytics industry has made a significant impact on

Indian companies, especially in their day-to-day operations. Our project has enabled the industry to completely automate some aspects of business. Data is considered to be the new oil, and industries.

- **References**

<https://github.com/aadarshprem82/Python/blob/b65fbb048ee17adae2e4ba243c8d814de4f28fec/hr%20database.xlsx>

- **Overview**

Data aggregation and analysis

Our project can assist in the collection of big data and solve major issues faced by big companies.

Discovering potential of untapped markets

India has numerous large untapped markets that can be identified, analyzed, and tapped into through data analytics.

Efficient allocation of resources

Our project can help a government, company or even an individual to allocate its resources efficiently.

Increased efficiency

Employment analysis helps companies boost efficiency.

## **Specific Requirement**

Once the components of the framework have been established can review the data that will need to be collected in order to measure the program indicators. This should also involve listing where data to be collected is being recorded and the questions, which they hope the project will respond to. As each question is reviewed, consider what data will address it. Questions may be listed in a table format alongside the specific variables being monitored to ensure there are no gaps in

the data. Once all questions and indicators have been linked in this way.

- **Functions**

Employment analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. This allow us to make informed decisions and to stop guessing

- **Logical Database Requirement**

A logical database is the collected information stored on multiple physical disk files and hard drives within a computer. This database provides a structure to house all the accumulated information within the device and determines the relationships between different types of files and programs.

A logical database can stretch over multiple physical hard disks information files. The data storage unit is still a single database for information retrieval purpose. To have a logical database, all given hard disks and information files must be accessible from a single source.

Our database can be seen in this link

<https://github.com/aadarshprem82/Python/blob/9e27462e078f1265fe2625c9c66b2b88008e9340/hrdatabase.xlsx>

- **Hardware requirement**

PC or Laptop with minimum 4 GB RAM and 256GB Storage.

- **Software requirement**

Anaconda Python, Jupyter notebook

- **Software System Attribute**

### Reliability

Identifies the degree to which different raters agree in the components of a target work role or job. It is typically assessed using Python and Pandas.

### Availability

This describes the process of identifying the number of members belonging to a minority or disadvantages groped who are qualified for any available position within the organization. Those are often candidates who suffer discrimination within a culture.

### Security

It is an approach to cybersecurity that uses data collection, data aggregation and analysis tools for threat detection and security monitoring. An organization that deploys security analytics tools can analyze security events to detect potential threats before they can negatively affect the company's infrastructure and bottom line.

This combines big data capabilities with threat intelligence to help detect, analyze and mitigate threats, as well as persistent cyber threats and targeted attacks from external bad actors.

### Maintainability

It is used to predict maintainability of systems and equipment of any type, including avionics, ground and shipboard electronics, mechanical equipment, etc.

Our project will receive a uniform checking for bug and problem, throughout the time it will maintain and be ready to use all the time.



## Portability

Portability is a characteristic attributed to a computer program if it can be used in an operating system other than the one in which it was created without requiring major rework. Porting is the task of doing any work necessary to make the computer program run in the new environment.

# Database Link

## **HR Database:**

<https://github.com/aadarshprem82/Python/blob/b65fbb048ee17adae2e4ba243c8d814de4f28fec/hr%20database.xlsx>

## **Python:**

<https://docs.python.org/3/>

## **Pandas:**

<https://pandas.pydata.org/docs/index.html>

# Testing of the Project

Testing is a method to check whether the actual software product matches expected requirements and to ensure that the software product is Defect free. It involves execution of software/system components using manual or automated tools to evaluate one or more properties of interest. The purpose of software testing is to identify errors, gaps or missing requirements in contrast to actual requirements.

## Types of Testing

Functional testing: - User Acceptance testing

Non-functional testing: - performance, Scalability, Usability

Maintenance testing: - maintenance

```
In [3]: lname = emp[emp.LAST_NAME == 'McEwen']
lname[['FIRST_NAME', 'LAST_NAME', "DEPARTMENT_ID"]]
```

```
Out[3]:
```

		FIRST_NAME	LAST_NAME	DEPARTMENT_ID	
	58	Allan	McEwen	80	
55	153	Christopher	Olsen	COLSEN	01
56	154	Nanette	Cambrault	NCAMBRAU	01
57	155	Oliver	Tuvault	OTUVAULT	01
58	156	Janette	King	JKING	01
59	157	Patrick	Sully	PSULLY	01
60	158	Allan	McEwen	AMCEWEN	01
61	159	Lindsey	Smith	LSMITH	01
62	160	Louise	Doran	LDORAN	01
63	161	Sarath	Sewall	SSEWALL	01
64	162	Clara	Vishney	CVISHNEY	01
65	163	Danielle	Greene	DGREENE	01
66	164	Mattea	Marvins	MMARVINS	01

# Security

Security describes measures at the application level that aim to prevent data or code within the app from being stolen or hijacked. It encompasses the security considerations that happen during application development and design, but it also involves systems and approaches to protect apps after they get deployed.

Security may include hardware, software, and procedures that identify or minimize security vulnerabilities. A router that prevents anyone from viewing a computer's IP address from the Internet is a form of hardware application security. But security measures at the application level are also typically built into the software, such as an application firewall that strictly defines what activities are allowed and prohibited. Procedures can entail things like an application security routine that includes protocols such as regular testing.

Different types of application security features include authentication, authorization, encryption, logging, and application security testing. Developers can also code applications to reduce security vulnerabilities.

# **Future Scope**

## **Work for Administrator**

This project is very useful for that person who handles all the critical roles in the company and is responsible for organizing and maintaining personnel records, updating internal databases, preparing HR documents, like employment contracts and new hire guides.

## **Associates**

This project is also useful for obtaining, recording, and interpreting human resources information within a company. They are tasked with managing corporation human resources records and assisting new employees with enrollment procedures.

## **Analytics**

By using this Project all the tasks related to analytics for an organization, they identify issues and work to solve them. Moreover, they are also responsible for creating organizations policies and objectives. Analyzing and evaluating data and reports are their key tasks.

# References

Reference is a value that enables a program to indirectly access a particular data, such as a variable's value or a record, in the computer's memory or in some other storage device. The reference is said to refer to the datum, and accessing the datum is called dereferencing the reference.

## **HR Database:**

<https://github.com/aadarshprem82/Python/blob/b65fbb048ee17adae2e4ba243c8d814de4f28fec/hr%20database.xlsx>.

Reference is emulated and expanded upon print reference forms including the dictionary, translation dictionary etc.

## **Python:**

<https://docs.python.org/3/>

## **Pandas:**

<https://pandas.pydata.org/docs/index.html>