System Analysis & Design

Problem Statement & Objectives: The primary problem addressed in this project is the high attrition rate in organizations, which affects workforce stability and operational costs. The objective is to analyze past data and predict employee turnover to help HR managers make data-driven decisions.

Use Case Diagram & Descriptions: The system involves HR managers as primary users, interacting with the predictive model through a Streamlit-based interface. The model processes employee data and provides attrition predictions.

Software Architecture: The project follows a modular design with the following components:

- **Frontend:** Streamlit-based web application.
- **Backend:** Python-based data processing and machine learning pipeline.
- **Database:** CSV-based storage with potential for database integration.
- **Machine Learning Models:** Implemented using Scikit-learn.

Deployment:

- The application is deployed using **Streamlit**, ensuring an easy-to-use web-based interface.
- Cloud hosting or local execution options are available for deployment.

Additional Deliverables:

- **Testing & Validation:** The project includes unit tests for data processing and model validation steps.
- **API Documentation (if applicable):** If an API is implemented in the future, documentation will be provided.
- **Deployment Strategy:** Streamlit-based deployment with GitHub integration for version control.

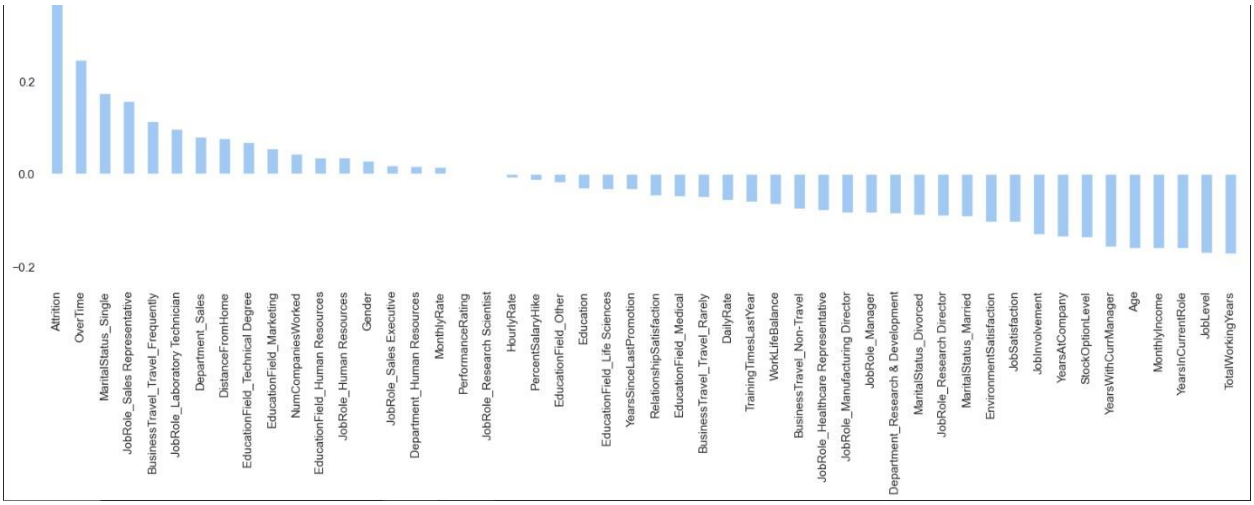
```
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...

```
df1.describe()
```

	Attrition	BusinessTravel	Department	EducationField	Gender	JobRole	MaritalStatus	Over18	OverTime
count	1470	1470	1470	1470	1470	1470	1470	1470	1470
unique	2	3	3	6	2	9	3	1	2
top	No	Travel_Rarely	Research & Development	Life Sciences	Male	Sales Executive	Married	Y	No
freq	1233	1043	961	606	882	326	673	1470	1054

The correlation between features and the targeted feature



The correlation between features (HeatMap)