CORPORATE EMPLOYEE ATTRITION ANALYSIS

PROJECT REPORT

SUBMITTED BY

KIRUTHIGA M

SUVETHA A R

SWETHA V

BHAVANI B

COMPUTER SCIENCE AND ENGINEERING

GRT INSTITUTE OF ENGINEERING AND TECHNOLOGY

TIRUTTANI

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The NalaiyaThiran opportunity I had with IBM in the Data Analytics domain was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual for being a part of it. I am also grateful to all the professionals who led me though this internship period.

I express my deepest thanks to my industry mentor Mr. Vinayagam for taking part in useful decision & giving necessary advices and guidance and arranged all facilities to make internship easier. I choose this moment to acknowledge his contribution gratefully.

I perceive as this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives.

Sincerely,

KIRUTHIGA M

SUVETHA A R

SWETHA V

BHAVANI B

OBJECTIVE

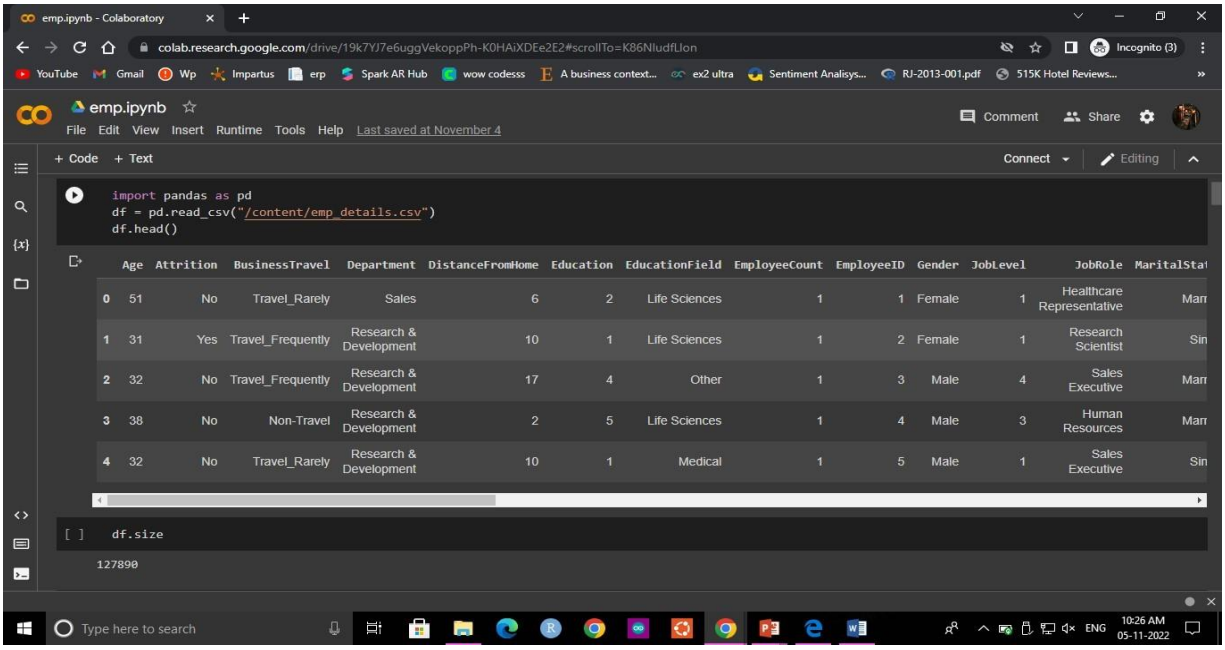
- The objective of this project is to predict the attrition rate for each employee, to find who's more likely to leave the organisation.
- It will help organization to find ways to prevent attrition or plan in advance the hiring of the new candidate.
- Attrition proves to be a costly and time-consuming problem for the organization and it also leads to the loss of probability.
- The scope of the project extends to companies to all industries

DESCRIPTION OF PROJECT

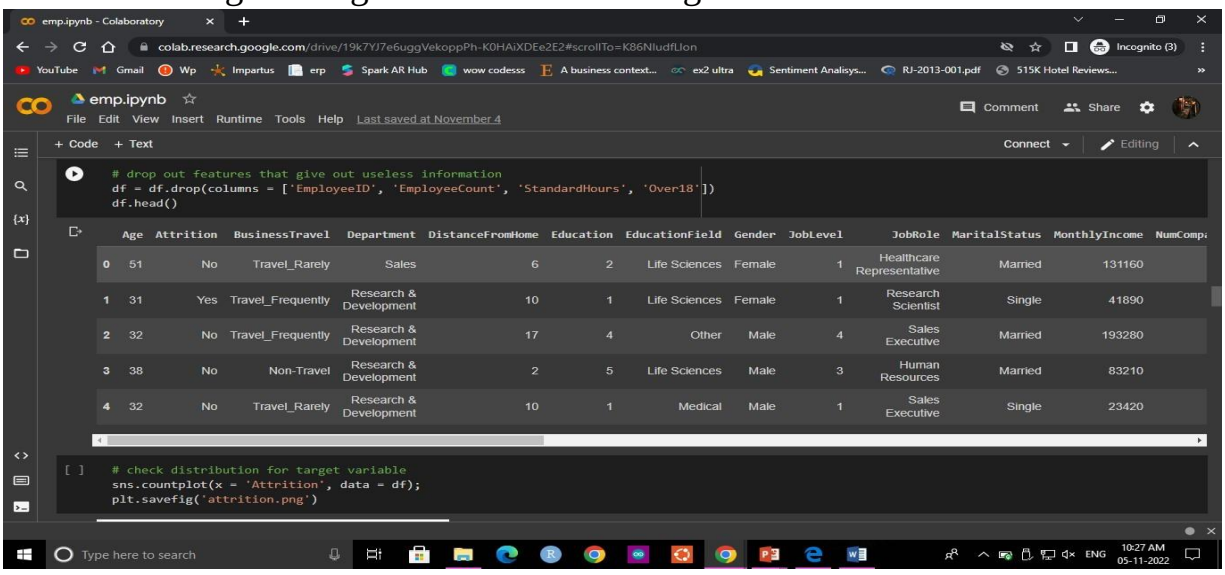
In this Project, it is required to clean and sanitize the dataset. Then, train the dataset to predict the attrition rate of the employees in an organization

VISUALIZATIONS

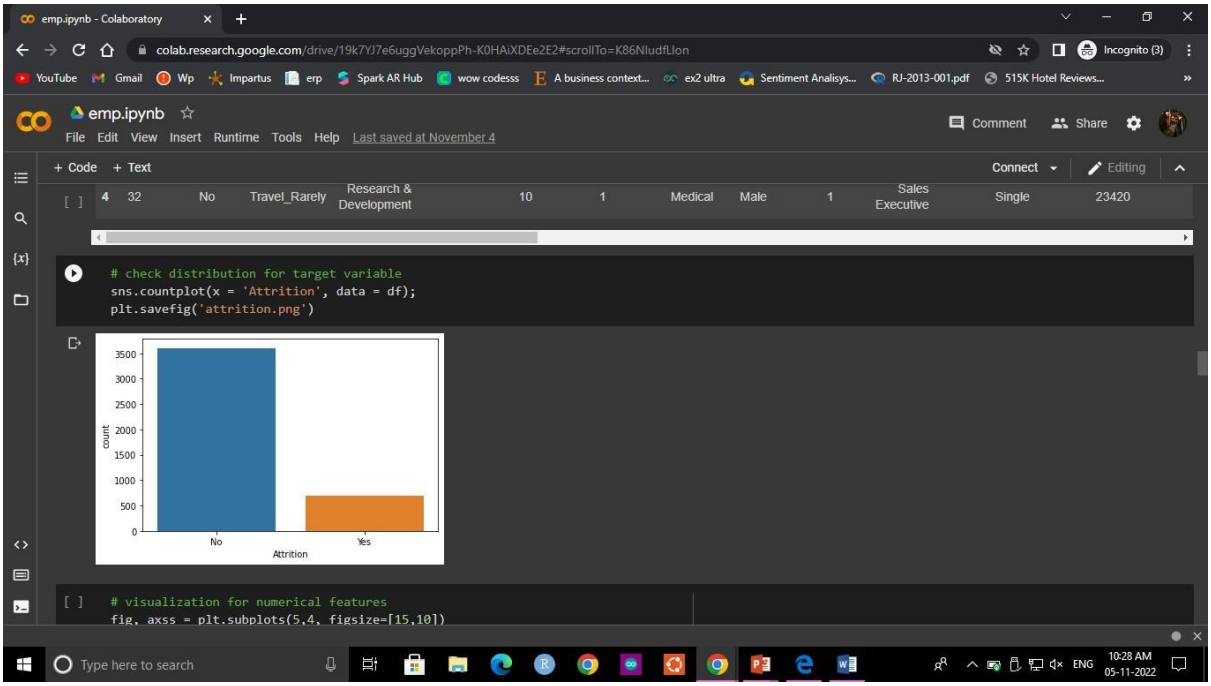
This figure shows the data frame (which isn't cleaned and sanitized as it has lots of null values)



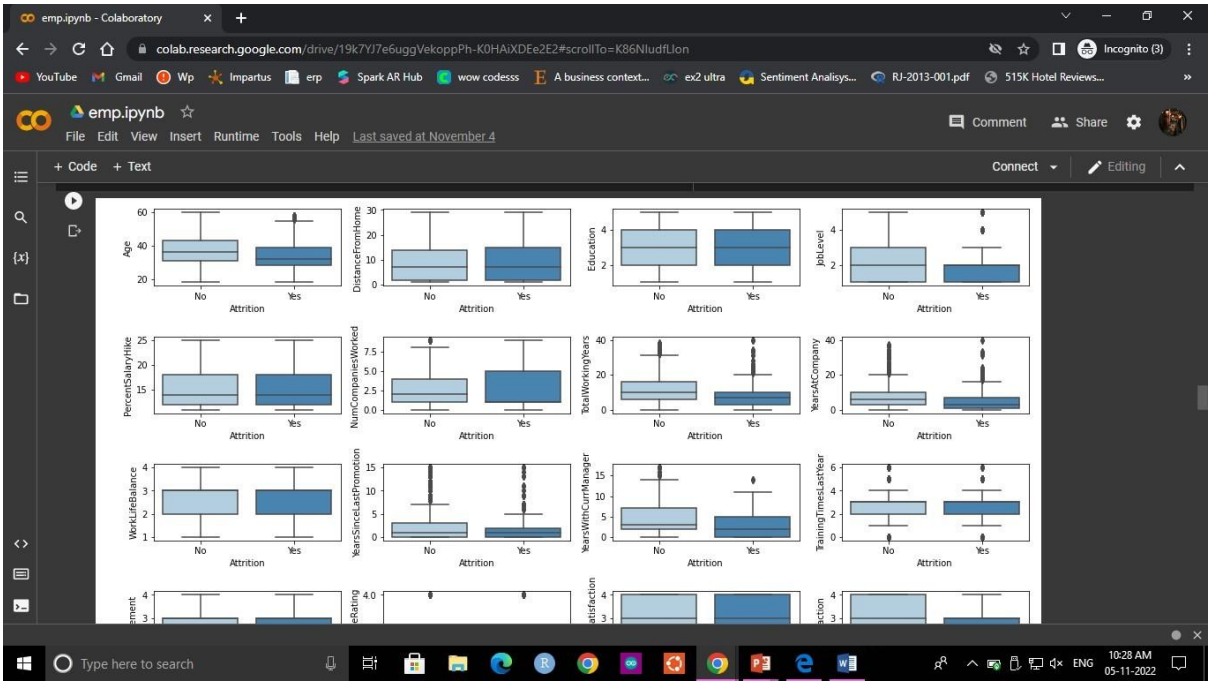
The below figure shows the dataset after removing unnecessary columns and the rows containing missing values and reordering the same.



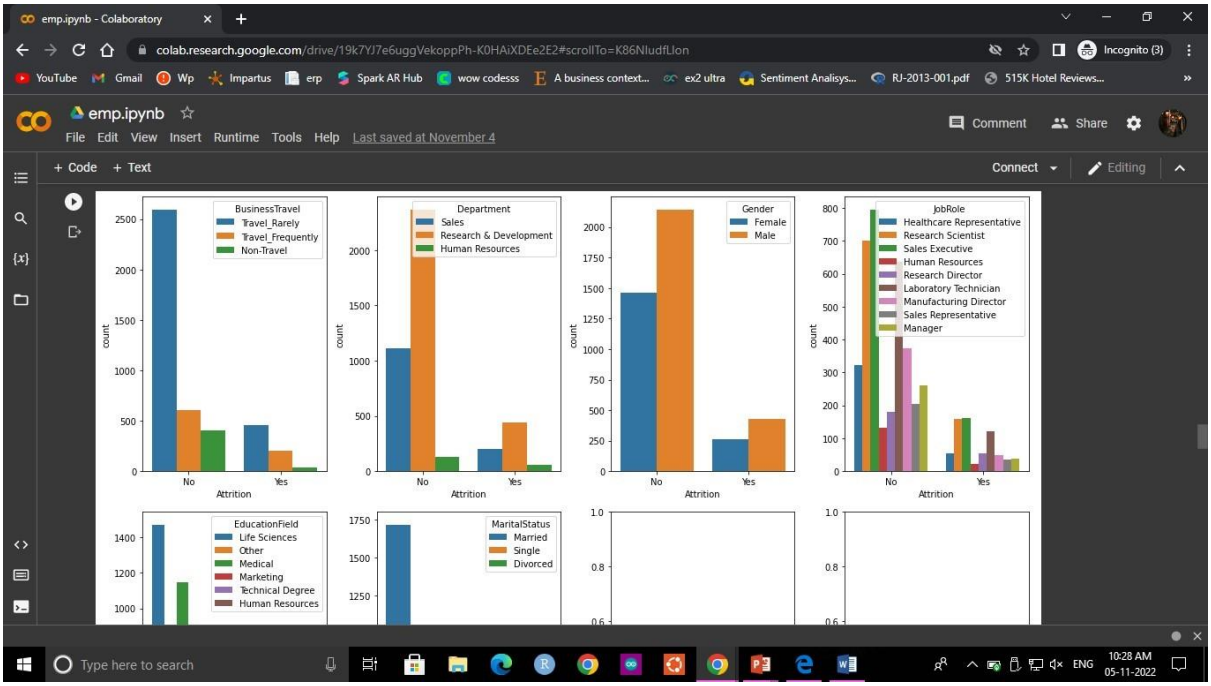
This figure shows the distribution of target variables



The visualisations of the numerical variables is



Similarly of the non-numerical variables



And then, we train the model and test for unknown data to predict the attrition rate. And the screenshot is attached herewith.

```
# Logistic Regression
start_time = time.time()
train_pred_log, acc_log, acc_cv_log = fit_ml_algo(LogisticRegression(), X_train,y_train, 10)
log_time = (time.time() - start_time)
print("Accuracy: %s" % acc_log)
print("Accuracy CV 10-Fold: %s" % acc_cv_log)
print("Running Time: %s" % datetime.timedelta(seconds=log_time))
```

```
Accuracy: 89.8
Accuracy CV 10-Fold: 88.63
Running Time: 0:00:01.753627
```

```
# SVC
start_time = time.time()
train_pred_svc, acc_svc, acc_cv_svc = fit_ml_algo(SVC(),X_train,y_train,10)
svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_svc)
print("Running Time: %s" % datetime.timedelta(seconds=svc_time))
```

```
Accuracy: 88.53
Accuracy CV 10-Fold: 85.91
Running Time: 0:00:00.497278
```

```
# Linear SVC
start_time = time.time()
train_pred_svc, acc_linear_svc, acc_cv_linear_svc = fit_ml_algo(LinearSVC(),X_train, y_train,10)
linear_svc_time = (time.time() - start_time)
print("Accuracy: %s" % acc_linear_svc)
print("Accuracy CV 10-Fold: %s" % acc_cv_linear_svc)
print("Running Time: %s" % datetime.timedelta(seconds=linear_svc_time))
```

```
Accuracy: 89.89
Accuracy CV 10-Fold: 88.73
Running Time: 0:00:01.055932
```

REFLECTIONS ON THE PROJECT

Our NalaiyaThiran project in collaboration with IBM and Tamil Nadu government has been the most rewarding and learning experiences we have had.With such empathetic, compassionate and supportive mentors, this experience has helped me achieve my goal of completing my project. Because of the techniques, we learned not only from my mentors and professors but from internet and books too.

We are confident that we will continue to grow and develop professionally and in my personal endeavours. Within my internship, there were two distinct learning experiences that stand out to me as the most influential aspects of my development this semester: community involvement in discussion forum and self-learning.

Throughout my project experience, we were able to develop and foster a truly positive and compassionate learning cum implementation environment, all through the support and mentorship of our mentors.

Through the application of time management, organization, discipline and consistent practice, our self-exploration and learning skills improved greatly. Additionally, my development both with the project we were given with and planning and implementing the same directly impacted our academic gain.

We are confident in our growth and development. We would not have the knowledge or skills we have today if it were not for our project experience with the industry mentor, college mentors and fellow interns.

CONCLUSION

On the whole, this project was a useful experience. We have gained new knowledge and skills we achieved several of my learning goals. We got insight into professional practice. We learned the different facets of working. We experienced that self-exploration, as in many organisations, is an important factor for the progress of projects. Related to our study we learned more about employee attrition rate prediction and the various approaches and algorithms to achieve the same. There is still a lot to discover and to improve. The methods used at the moment are still not standardized and a consistent method is in development. Furthermore we have experienced that it is of importance of each strategy and how other one is better than the current algorithm and in which application. We found that the internship is not one sided, but it is a way of sharing knowledge, ideas and opinions and implementing the same to get results. The internship was also good to find out what our strengths and weaknesses are. This helped me to define what skills and knowledge. We believe that our time spent in learning and surfing regarding various algorithms and the mathematics behind was well worth it and contributed to finding an acceptable solution to build a model and predict the employee's attrition rate. Two main things that we've learned the importance of time-management skills and self-motivation. At last this project has gave us new insights and motivation to pursue a career in machine learning domain.

Link to code and executable file

https://github.com/IBM-EPBL/IBM-Project-29002-1660119975