Problem:

> According to the 2014 Training Industry Report from Training magazine, the annual training budgets of U.S. small businesses totaled an average of $308,000 in 2014, with retail and wholesale spending more than half a million dollars. Approximately $1,200 was spent per employee.
>
> - **Forbes**

Costs appear not just in recruiting and onboarding, but costs other employees time by bringing new hires up to speed.

Context:

It's important to identify which employees are at risk for attrition and more importantly, why they quit and how to retain workers.

We can use this data to reduce pain points to create better environments for employees, reducing administrative costs of onboarding, and freeing up time supervisors/mentors use to bring a new hire up to speed, which could be used working on projects and developing strategy.

Criteria for success:

Predicting employees who will quit (minority class) with an F1 score of above 50% and identifying the contributing factors to their attrition.

Constraints:

The data source from IBM is a fictional dataset that might not be representative of real work working data.

Additionally, some features of the dataset have unclear definitions or meanings, making them harder to interpret if they explain significant variance in the model.

The data set also does not contain specific datetime data, only length of employment in years, possibly leaving out important seasonal attrition.

Data Sources:

IBM HR Analytics Employee Attrition and Performance (fictional data set created by IBM):
https://www.kaggle.com/uniabhi/ibm-hr-analytics-employee-attrition-performance