**Project Option 1:** [**https://www.kaggle.com/esmaeil391/ibm-hr**](https://www.kaggle.com/esmaeil391/ibm-hr)

1. **Draft a well-formed problem statement relevant to a business problem affecting your team, division, or organization.**

Working at a training centre for HR professionals, HR analytics is something frequently spoke at a superficial level, but not looked into due to its technical aspects. During work, I have also encountered many employers who have said that employees are hard to keep. Hence, I hope to use the IBM HR dataset specially created for IBM Watson to determine attrition of employees to create a “model” to identify what makes people quit.

1. **Include the following elements:**
   * **Hypothesis/assumptions**

People quit for money.

* + **Goals and success metrics**

Easy: Predict quitters with high accuracy.

Difficult: Reduce attrition rate by “adjusting” data in dataset after prediction.

* + **Risks or limitations**

Dataset has been around for some time, so it has been overdone with IBM’s original url leading to other pages.

Need to change some of the int in dataset back to string?

Imbalance data

1. **Identify at least one relevant internal dataset and confirm that you have (or can get) the right access permissions.**

Public data created with fictional employees. <https://www.kaggle.com/esmaeil391/ibm-hr>

**Backup Choice:** [**https://www.kaggle.com/kemical/kickstarter-projects/**](https://www.kaggle.com/kemical/kickstarter-projects/)

1. **Draft a well-formed problem statement relevant to a business problem affecting your team, division, or organization.**

Getting out of my depressing job, I need a backup plan. I need to know what kind of project is able to garner the most investments. Hence, I look at Kickstarter projects success rates. It can also become a model for anyone looking to get money from anyone.

1. **Include the following elements:**
   * **Hypothesis/assumptions**

The longer the time, the more chance the project is successful.

The lesser the money needed, the more successful the project.

* + **Goals and success metrics**

Easy: Predict how accurately how a Kickstarter can be successful.

* + **Risks or limitations**

It is a big dataset in terms of row. A parameter may need to be set even before the project.

Feature columns might need to be re-looked at to provide more meaningful data.

True feature affecting the target might be from beyond the features in dataset.

1. **Identify at least one relevant internal dataset and confirm that you have (or can get) the right access permissions.**

Public data. <https://www.kaggle.com/kemical/kickstarter-projects/>