Type your notes here and we will discuss them in the meeting

* any anomalies (e.g. some date values within company\_size column, data type, missing values, outliers...)
* Questions for EDA

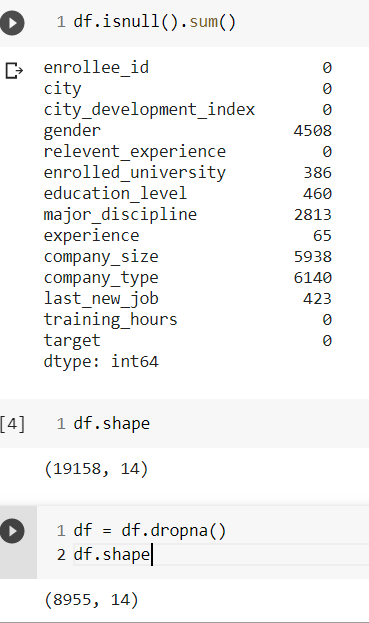
**Summary:** Fill in the missing values with all “Unknown” in the **Visualization Process**

**Add 2-3 sentences** to describe what you found out in the dataset

Phase1:

* 2 general questions + 3 features specific Questions
* Story lines
* **Meeting: 11:30am saturday in-perosn**

Columns with missing values:

* Gender -- No way to backfill data for ML model. () we can just get rid of those rows?
* Enrolled\_university
* Education\_level
* Major\_discipline
* Company\_size
* company\_type
* Chris
  + Training hours with target
* Jacinto
  + Company size, company type
* Aash
  + CDI, Gender
* Caroline L
  + Enrolled\_university, education\_level, major\_discipline
* Caroline G - Relevant\_experience, Experience, last\_new\_job
* Caroline L
  + Missing values in Gender Column ---> Delete from 19,159 -> 14,651 rows
  + Major\_discipline ----> missing values filled with No Major
  + Missing values in Experience column
  + 5 rows showing date values within company\_size column,
  + Missing value in company\_type ---> Company N/A
* Caroline G
  + ??
* Jacinto
  + Missing values in gender(24%), enrolled\_university, education\_level, major\_discipline(15%), company\_size, company\_type. Relevent\_experience and experience may be interesting variables to look at...the more experience you have, the pickier you are at job selection and dont want to hop around from job to job. Company\_type is interesting too.. For industries with a high learning curve , people tend to stay for a few years and then go somewhere else.
  + Columns picked: Company\_size, company\_type, target
* Chris
  + Quite a bit of missing values in the company size and type columns
  + Gender predictor: Are men or women more likely to look for new employment or continue to work for the company after training
* Aash
  + 19K rows. With a lot of missing values. If all missing values are dropped final dataset is 8K. Should we drop missing values or fill it with averages??
  + EDA
    - Gender vs target (More men or women likely to switch)
    - Relevant Experience vs Target (PPL who have the relevant skillset vs who dont, and likeness to switch)
    - Education level vs Target
    - Company SIze, Company Type, last new job vs Target.
    - Training hours vs target.
* 

**780 Project Process**

* **~~Step 1: Problem Formulation~~**
  + **~~Build predictive model to determine whether a candidate is looking for a new job or not.~~**
* **Step 3: Exploratory Data Analysis** 
  + **Questions**
    - **Analyzing how each variable is related to the target variable**
    - **Cleaning + Visualization**
* **Step 4: Predictive Model (classification problem)**
  + **Data Preparation and Preprocessing**
  + **Feature Selection**
  + **Model Development**
  + **Model Evaluation**
* **Step 5: Results (Recommendations)**
  + **What benefit will it bring?**
    - **It helps to save cost of human resources**
  + **feedback of the company’s working environment and guidance in the future recruiting**