**HR Analytics Case Study**

1. Problem Definition

Every year a lot of companies hire a number of employees. The companies invest time and money in training those employees, not just this but there are training programs within the companies for their existing employees as well. The aim of these programs is to increase the effectiveness of their employees. But where HR Analytics fit in this? and is it just about improving the performance of employees?

**HR Analytics**

Human resource analytics (HR analytics) is a subset of analytics that involves applying analytic techniques to an organization's human resource department in the hopes of enhancing employee performance and thereby increasing return on investment. HR analytics is more than just collecting data on employee productivity. Rather, it seeks to provide information about each step.

**Attrition in HR**

Attracting human resources means the gradual loss of employees' overtime work. In general, companies face the problem of relatively high staff turnover. HR professionals often take the lead in developing a company's compensation programs, work culture, and incentive systems that help organizations retain the best employees.

How does Attrition affect companies? and how does HR Analytics help in analysing attrition?

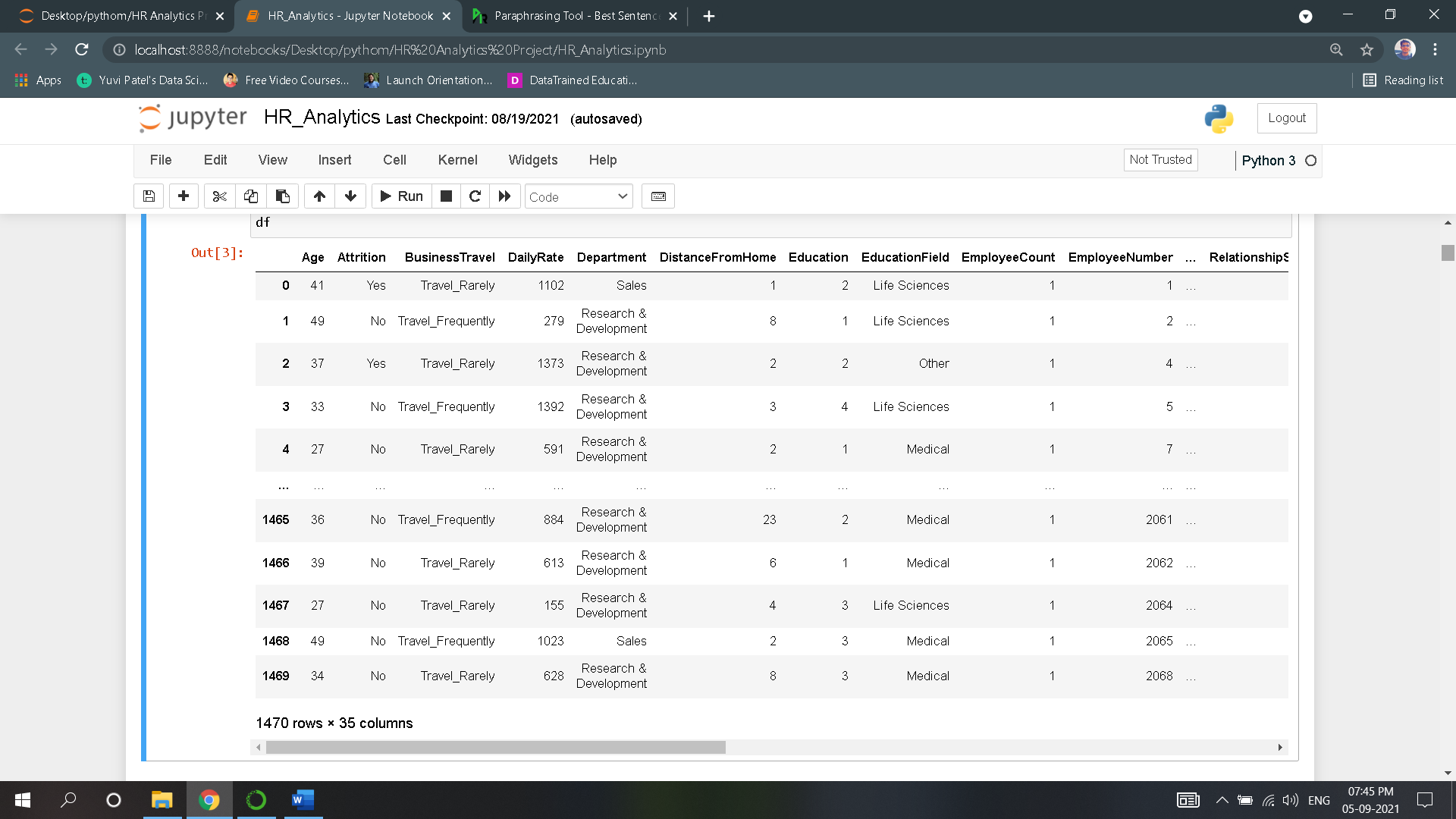
**Attrition affecting Companies**

The main concern with leaving many employees is the cost to the organization. Publishing, hiring, paperwork and training new hires are some of the most common costs of losing and replacing employees. In addition, frequent membership changes prevent your organization from increasing its collective knowledge and experience over time. This is especially important if your business is customer-centric, as customers often enjoy interacting with people they know. There will be more errors and crashes if you constantly have new workers.

2. Data Analysis

2.1 Data Import:

* The dataset is in csv format, we shall import the dataset using ‘read\_csv’ function.
* Once the dataset is imported and converted into a data frame, store the data frame and print it, to analyse the datapoints in rows and columns.



* Our requirement is to predict Attrition with the help of several different attributes.
* Each column in the Data frame is as follows:

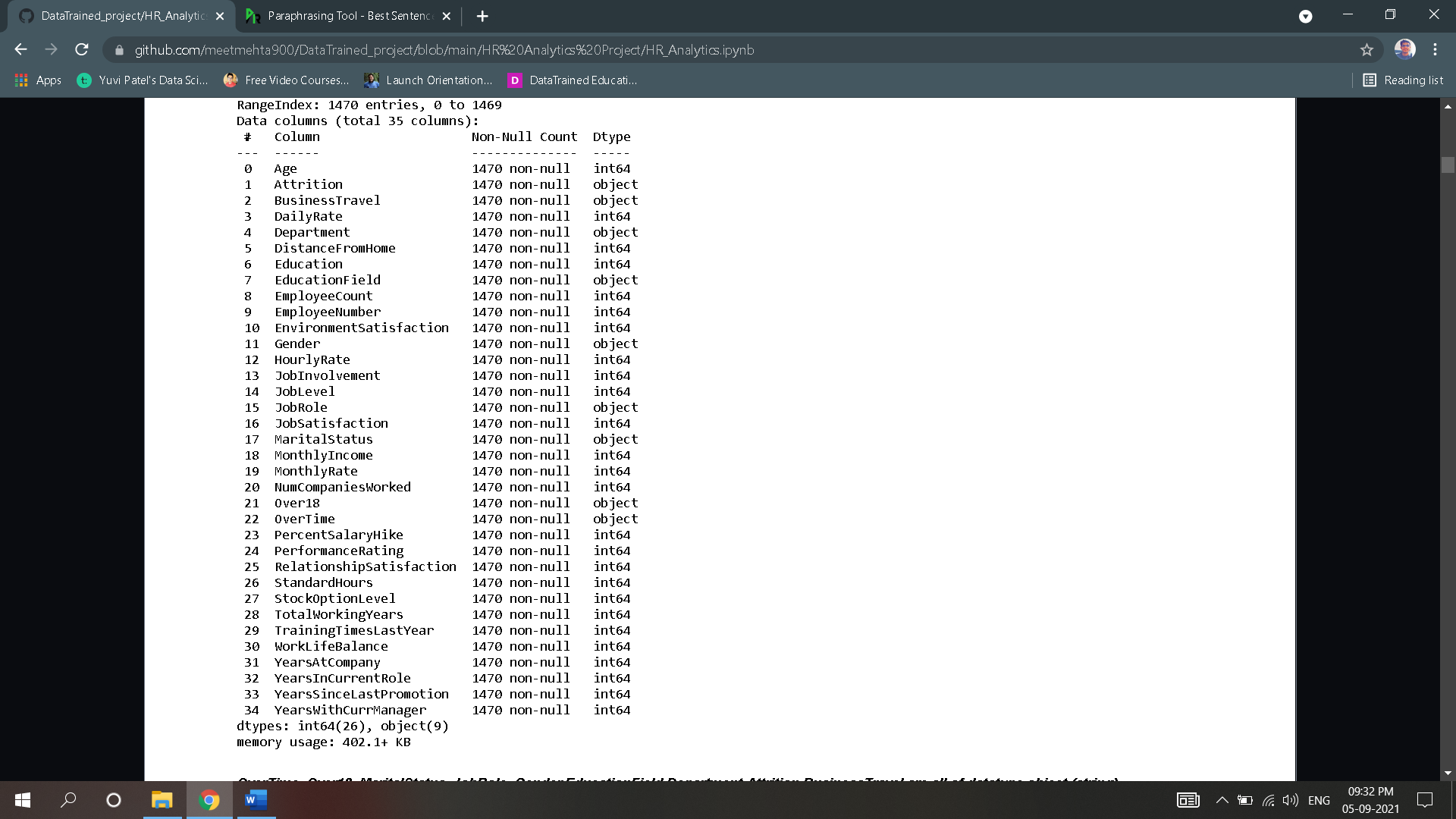
Output/Target Variable:

* Attrition: whether employees get fired or not

Input/Independent Variable:

* Age
* Business Travel
* Daily Rate
* Department
* Distance from Home
* Education
* Education Field
* Employee Count
* Employee Number
* Environment Satisfaction
* Gender
* Hourly Rate
* Job Involvement
* Job Level
* Job Role
* Job Satisfaction
* Marital Status
* Monthly Income
* Monthly Rate
* No. of Companies Worked
* Over18
* Overtime,
* Percent Salary Hike
* Performance Rating
* Relationship Satisfaction
* Standard Hours
* Stock Option Level
* Total Working Years
* Training Times Last Year
* Work-Life Balance
* Years at Company
* Years in Current Role
* Years Since Last Promotion
* Years with Current Manager

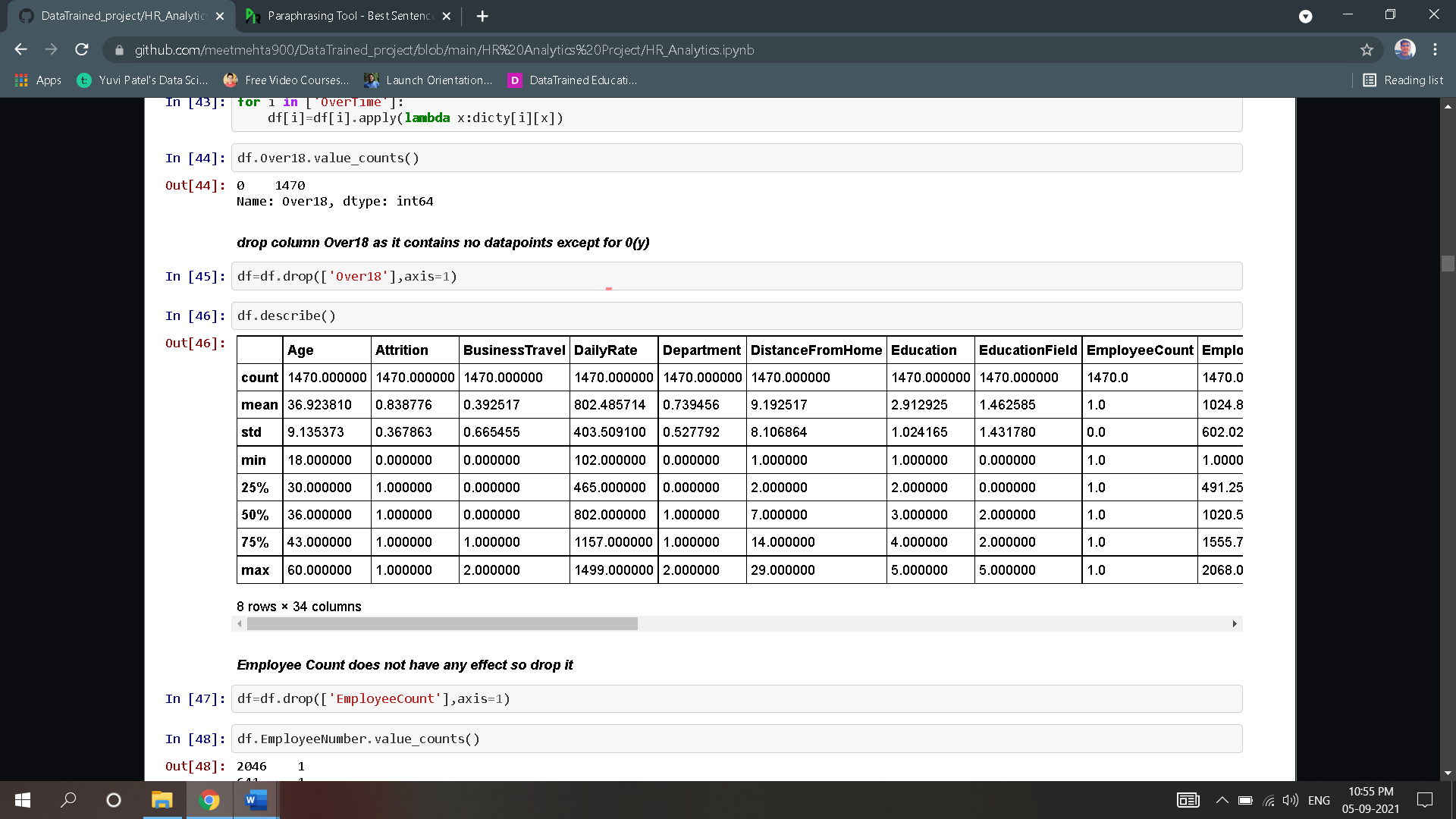
2.2 Data Analysis



* After executing the ‘info’ command we obtain the above output and following is our observation:
* Number of rows in dataset are:1470
* Number of columns in dataset are:35
* Dataset does not contain any null values.
* There are 26 numerical columns and 9 categorical columns which includes target variable.

2.2.1. Data Preparation

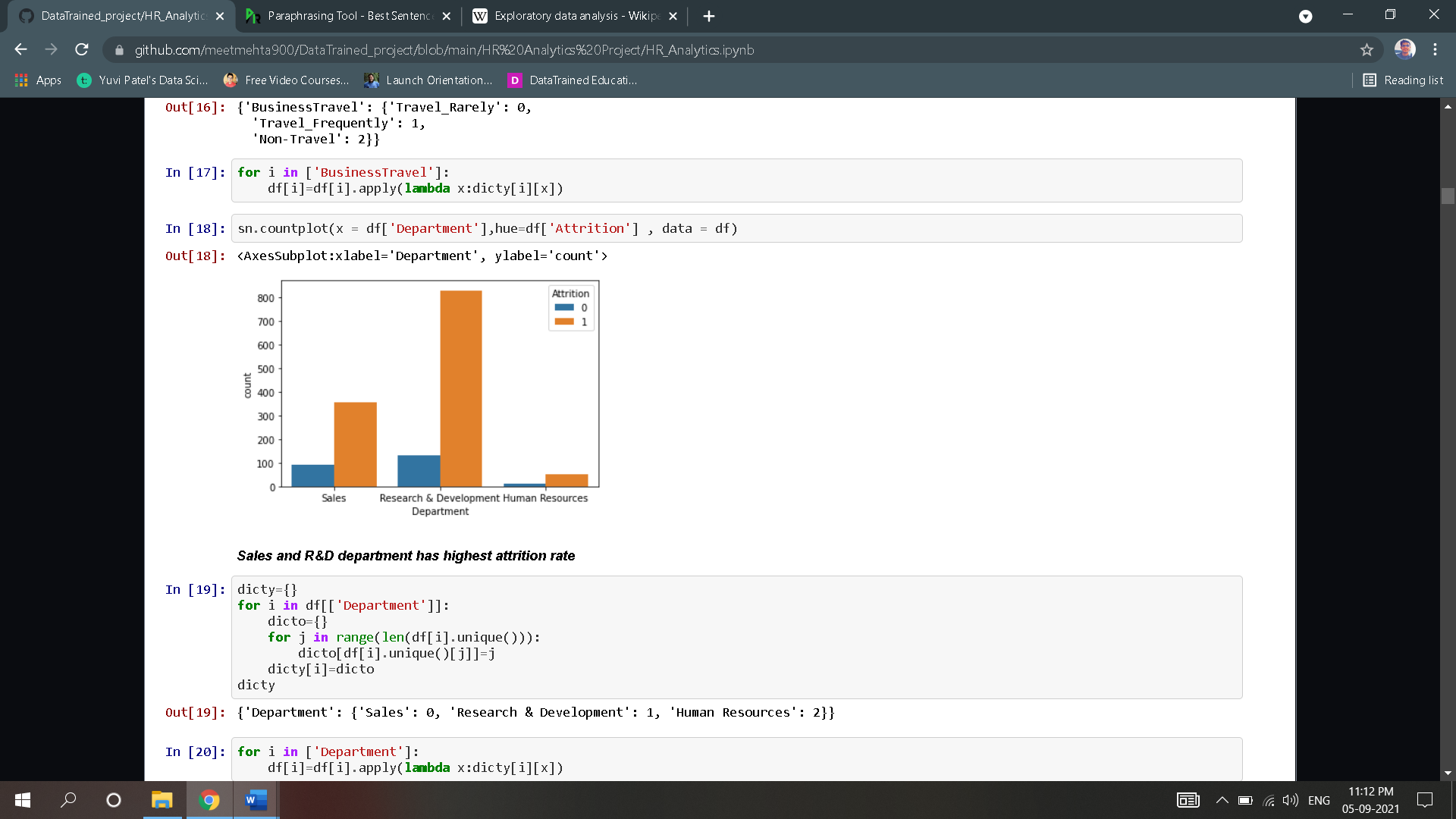
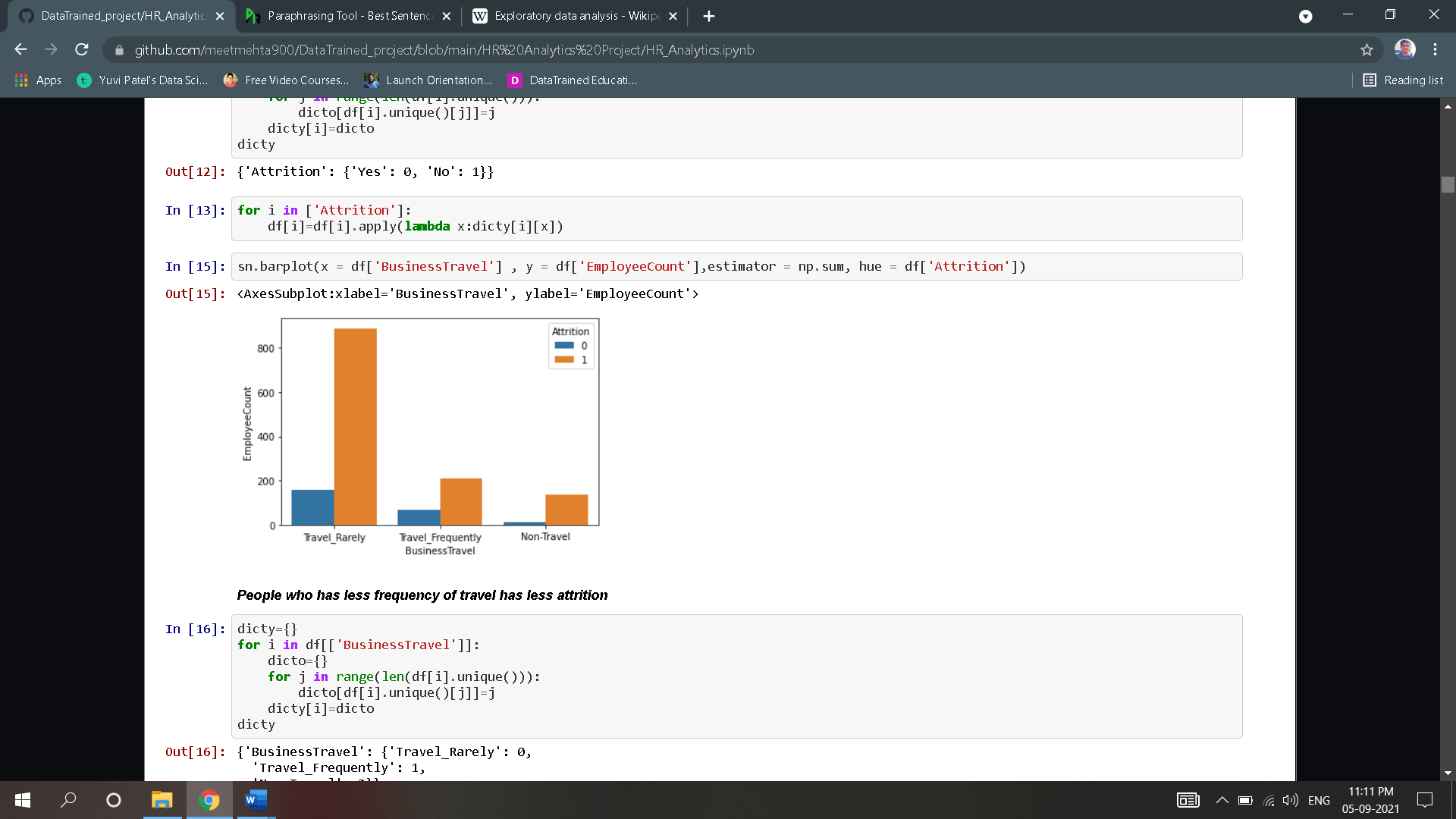
* Converting Categorical Columns into Numeric type as model which are about to build does not accept categorical values.
* For we need to know unique values of each categorical columns.
* {'Attrition': {'Yes': 0, 'No': 1}}
* {'BusinessTravel': {'Travel\_Rarely': 0,'Travel\_Frequently': 1,'Non-Travel': 2}}
* {'Department': {'Sales': 0, 'Research & Development': 1, 'Human Resources': 2}}
* {'EducationField': {'Life Sciences': 0,'Other': 1,'Medical': 2,'Marketing': 3,'Technical Degree': 4,'Human Resources': 5}}
* {'Gender': {'Female': 0, 'Male': 1}}
* {'JobRole': {'Sales Executive': 0,'Research Scientist': 1,'Laboratory Technician': 2,'Manufacturing Director': 3,'Healthcare Representative': 4,'Manager': 5, 'Sales Representative': 6,'Research Director': 7,'Human Resources': 8}}
* {'MaritalStatus': {'Single': 0, 'Married': 1, 'Divorced': 2}}
* {'Over18': {'Y': 0}}
* {'OverTime': {'Yes': 0, 'No': 1}}
* Now this Unique value will be converted to numerical type by using **Label encoder.**
* Now that all columns are converted to numeric type, we can have brief **Statistical Summary** containing important metrics such as Mean, Standard deviation, Minimum value, Maximum value of each column are obtained using describe function.
* These metrics gives us information about the possibility of outliers present in column, whether all values are balanced or not.



* Upon the above summary, there is huge variance in quartile range of many columns and standard is high comparatively which means there outliers present in columns which needs to taken care of in further analysis.

3.Exploratory Data Analysis

* Exploratory data analysis is an approach of analysing [data sets](https://en.wikipedia.org/wiki/Data_set) to summarize their main characteristics, often using [statistical graphics](https://en.wikipedia.org/wiki/Statistical_graphics) and other [data visualization](https://en.wikipedia.org/wiki/Data_visualization) methods.

##### Upon above plots, we can see that, People who has less frequency of travel has less attrition

##### Sales and R&D department has highest attrition rate.