A sample data file from IBM was used for analysis. Which contained approximately 7000 telecom customers with 20 attributes linked to users. The file name is BDA601_Assessment 2_Telco-Customer-Churn_downloaded 05082020.csv, which is attached to the paper. Four folders were created for each processing step. The first folder comprises the construction of the data set, the second folder contains the development of the data model, the third folder contains this report in PDF format and the fourth folder contains the interpretation of the analysis of the 'Churn' attribute.

Firstly, in folder Task1, with the original file, command lines were generated in Python language for the removal of 5 parameters that were not necessary for the deeper evaluation of the data. The following parameters were excluded: Monthly Fees, OnlineSecurity, StreamingTV, InternetService and Partner. As I used a language which only provided one exclusion of attributes at a time, it was necessary to create 5 data files. Different files were created to be able to follow the step by step of each process. After the exclusion of the last parameter, the data file new_data4.csv was generated. The database of new_data4.csv contains 16 attributes and 7,043 observations.

With the new database created, we set off to analyse the attributes, where the attribute 'Churn' will be our analysis target. In the Task2 folder, the database generated in the previous task was placed so that the data model could be developed. The whole process of transforming the information contained in the attributes into graphics and mathematical data was demonstrated in Python language for the creation of the following parameters. Presenting a data visualisation in an appropriate way and descriptive statistics. With this data, you can develop a 'decision tree' model to predict customer churn.

With the 'decision tree' model ready, the analysis of the attribute 'Churn' was interpreted. With an effectiveness of its churn analysis: a percentage of time from 70% to 30% was used.

All the processes are shown in the attached files.