Citizens & Media Analysis

By Gerard Utoware

Table of Contents

Overview of the tool used

The main advantage of python in comparison to other data analytic tools such as excel, would be Python's ability to handle large volumes of data without hindering productivity. Running scripts using the libraries on allows for more automation and ease of obtaining analysed data. With this task, the data was imported from python which ensures that data will not be lost or tampered with while performing data analysis. After defining the columns from the imported data, it was less tedious to run scripts than to perform individual analysis on excel.

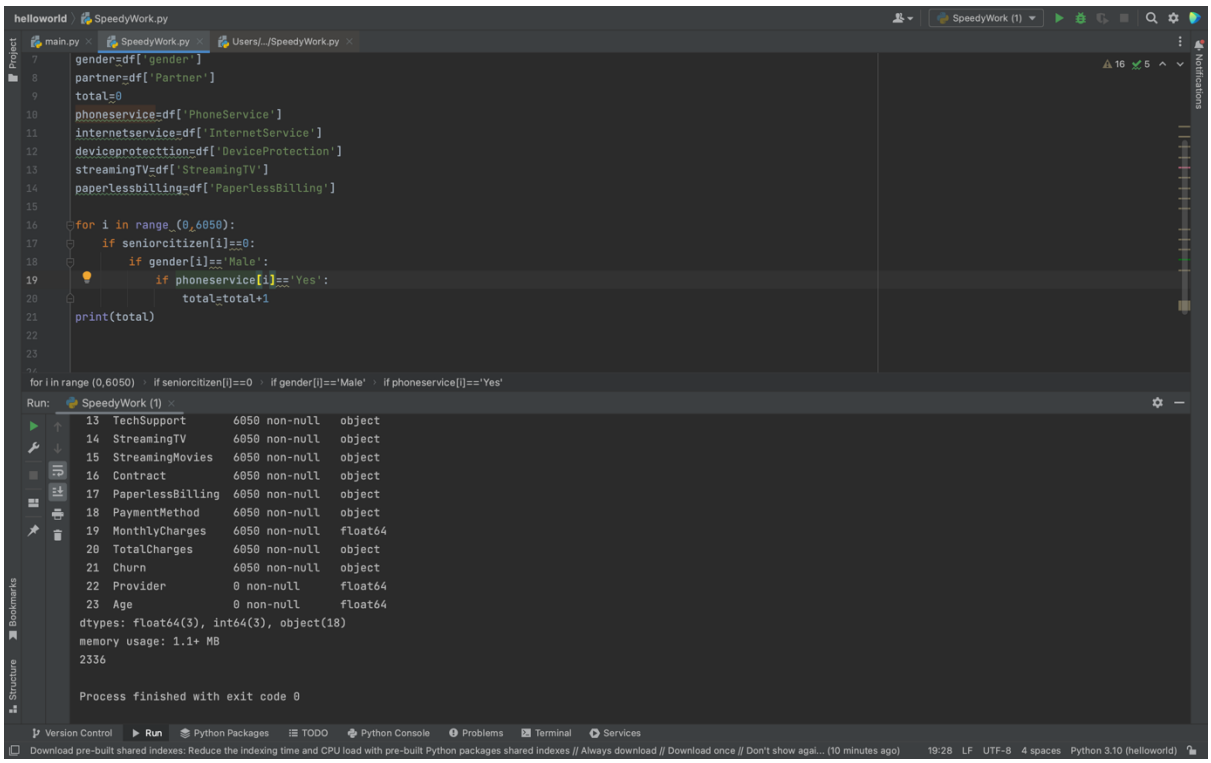
Overview of the libraries used

Numpy, pandas and matplotlib were used in the script. All three python libraries allow different functions to be performed. Pandas allows data analysis scripts to be conducted on python. It utilises two other libraries, being numpy and matplotlib. Matplotlib allows for data visualisation through a variety of charts such as bar and pie charts. Numpy permits mathematical operations in python. The combination of these three libraries is what allowed the following data and charts to be formed and analysed.

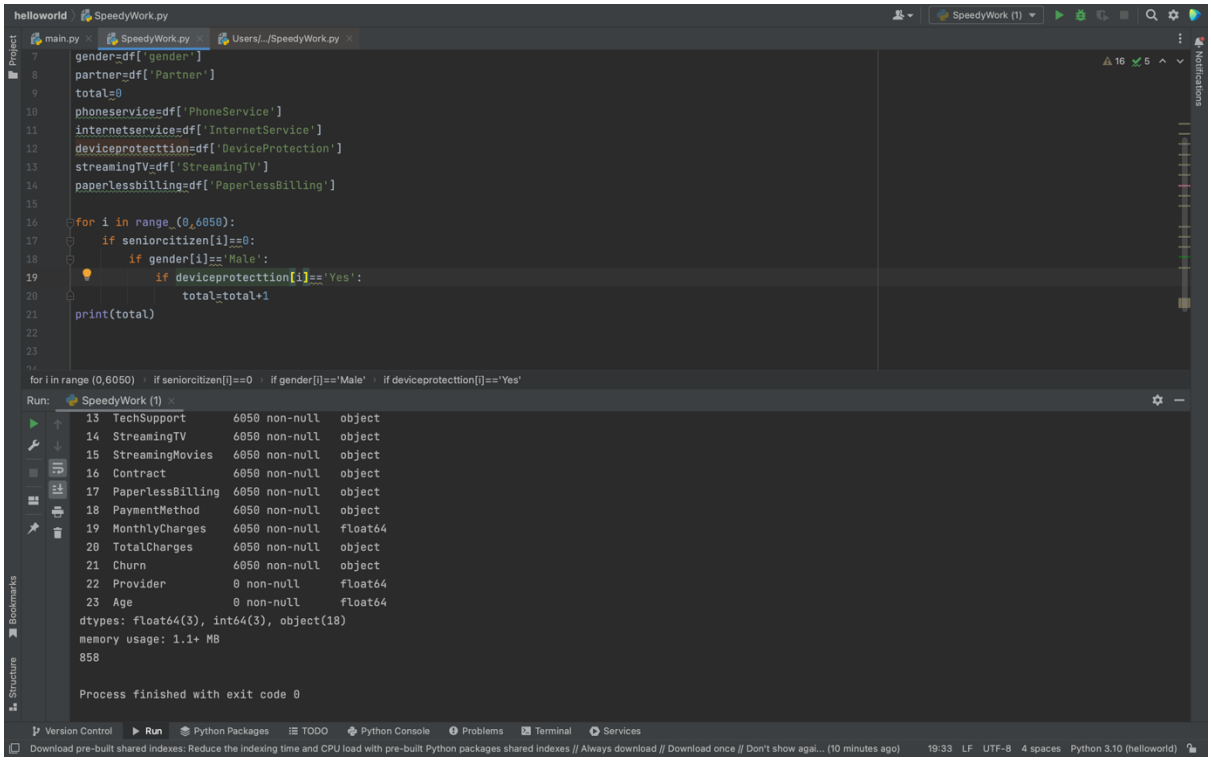
Overview of the dataset

Overall, the data suggests that there is a shift towards more modern uses of technology amongst the population. In regard to paperless billing, although most people have switched to paperless billing, approximately 40.9% of people still prefer to have their bills mailed to them. Senior citizens in the data appear to have a higher proportion of paperless billing as 748 senior citizens have paperless billing compared to 233 who do not. Gender did not seem to create a significant difference when comparing the same criteria. For example, when looking at non senior citizens, there was no significant difference between the number of males who had DSL or fiber optic (2591) or the females who had DSL or fiber optic (2478) as their internet service. Even when comparing non-Senior citizens with and without partners, the numbers did not deviate too much. Suggesting that both marital status and gender did not play a role in the likelihood of people having internet service, phone protection, streaming TV etc.

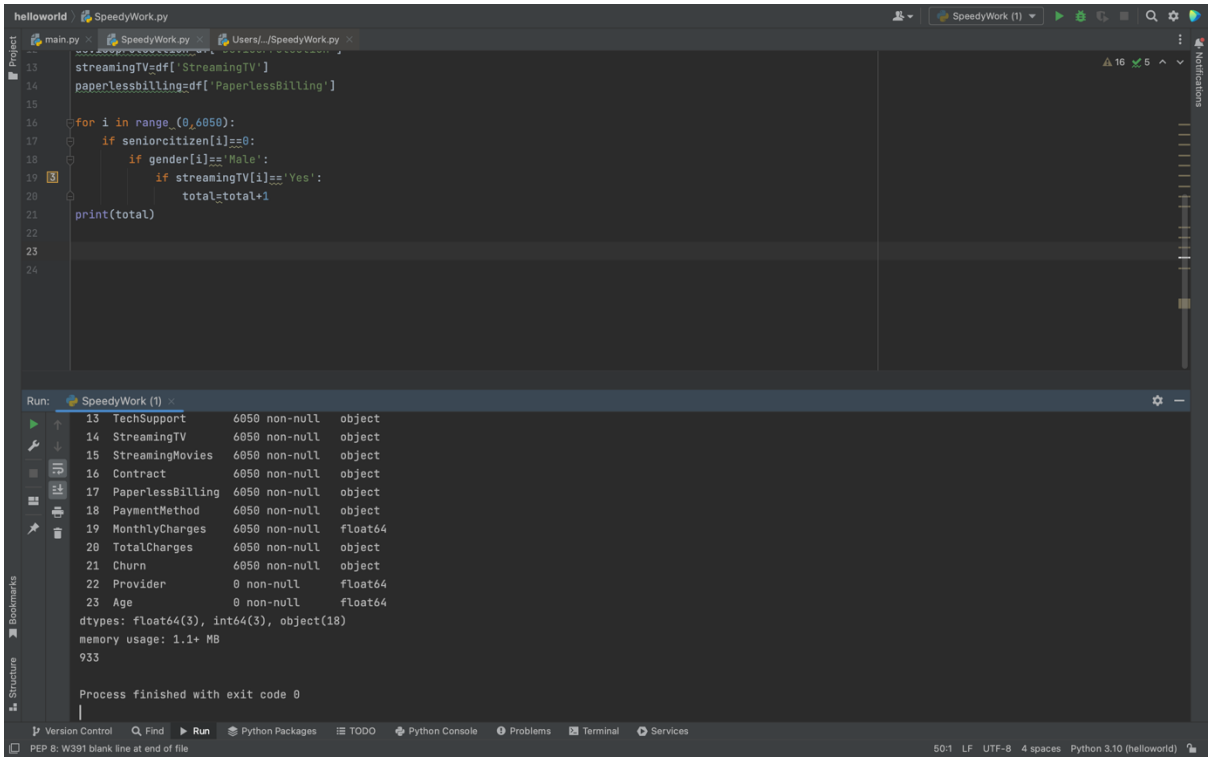
Q1



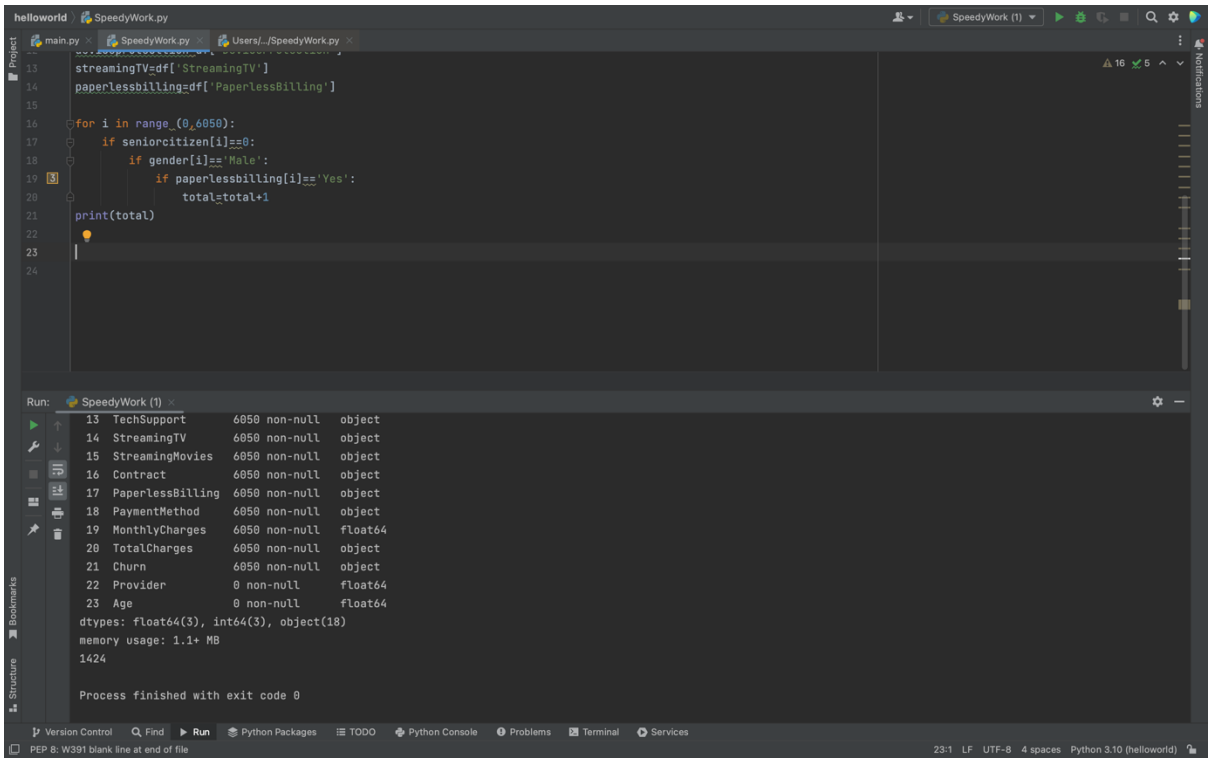
There are 2336 non-Senior Citizen Males that have phone service.



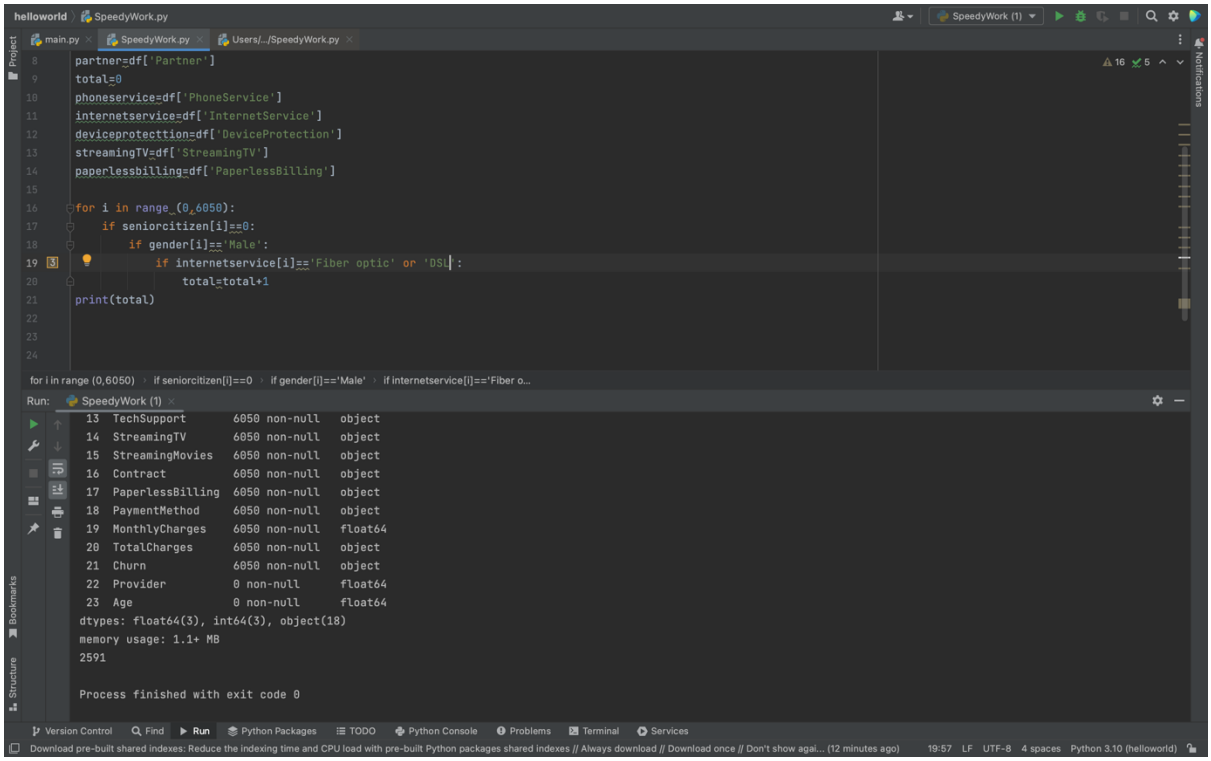
There are 858 non-Senior Citizen Males that have device protection.



There are 933 non-Senior Citizen Males that have a streaming TV.

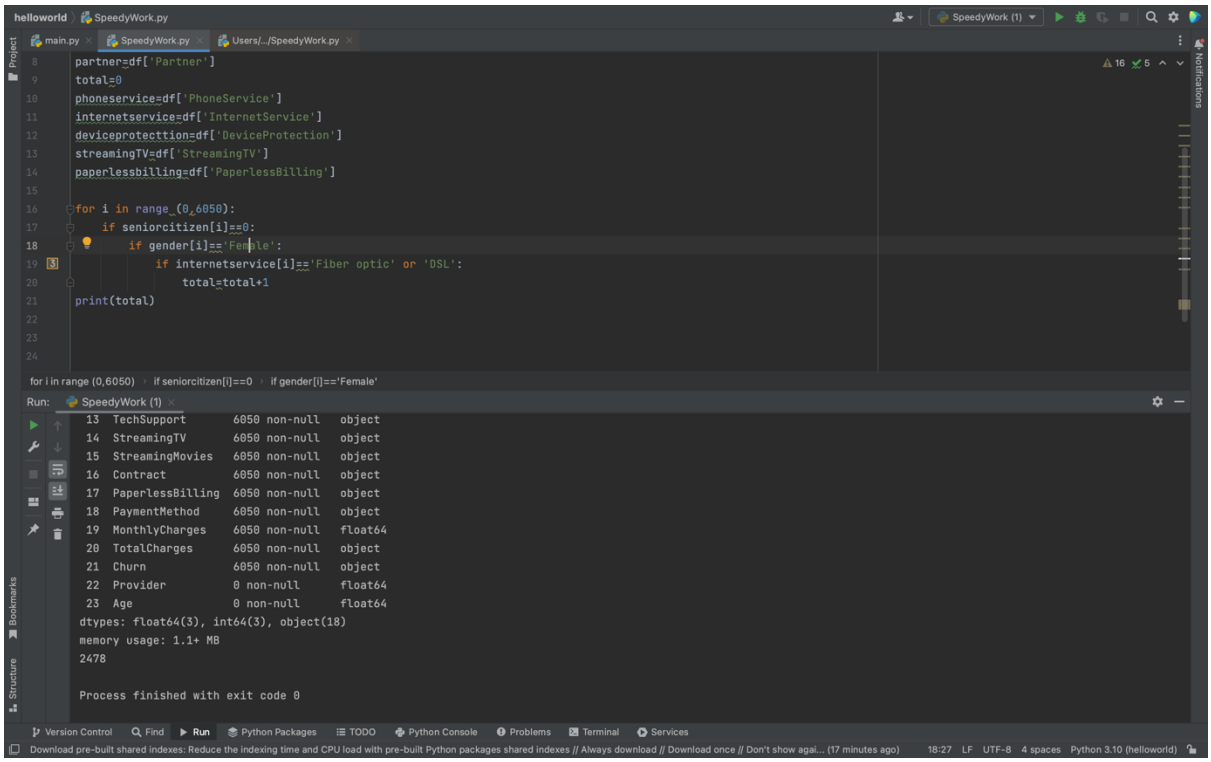


There are 1424 non-Senior Citizen Males that use Paperless billing.

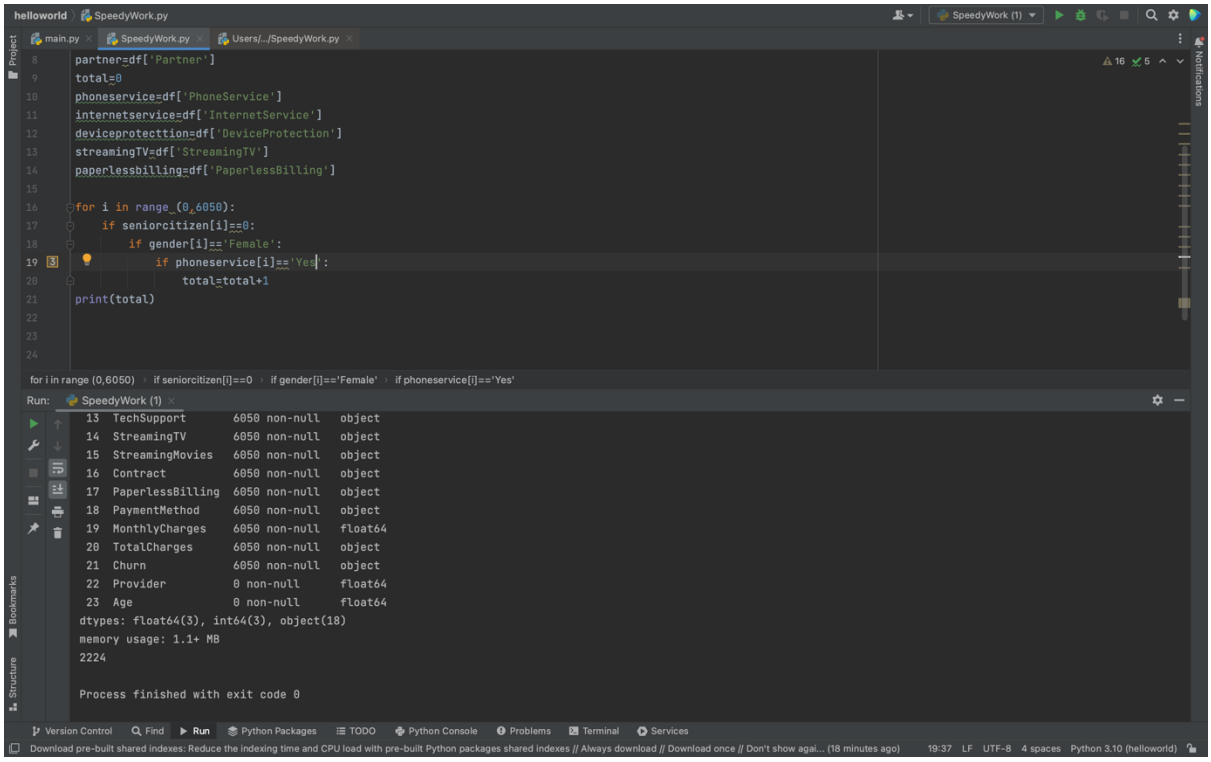


There are 2591 non-Senior Citizen Males that have either DSL or Fiber Optic Internet service.

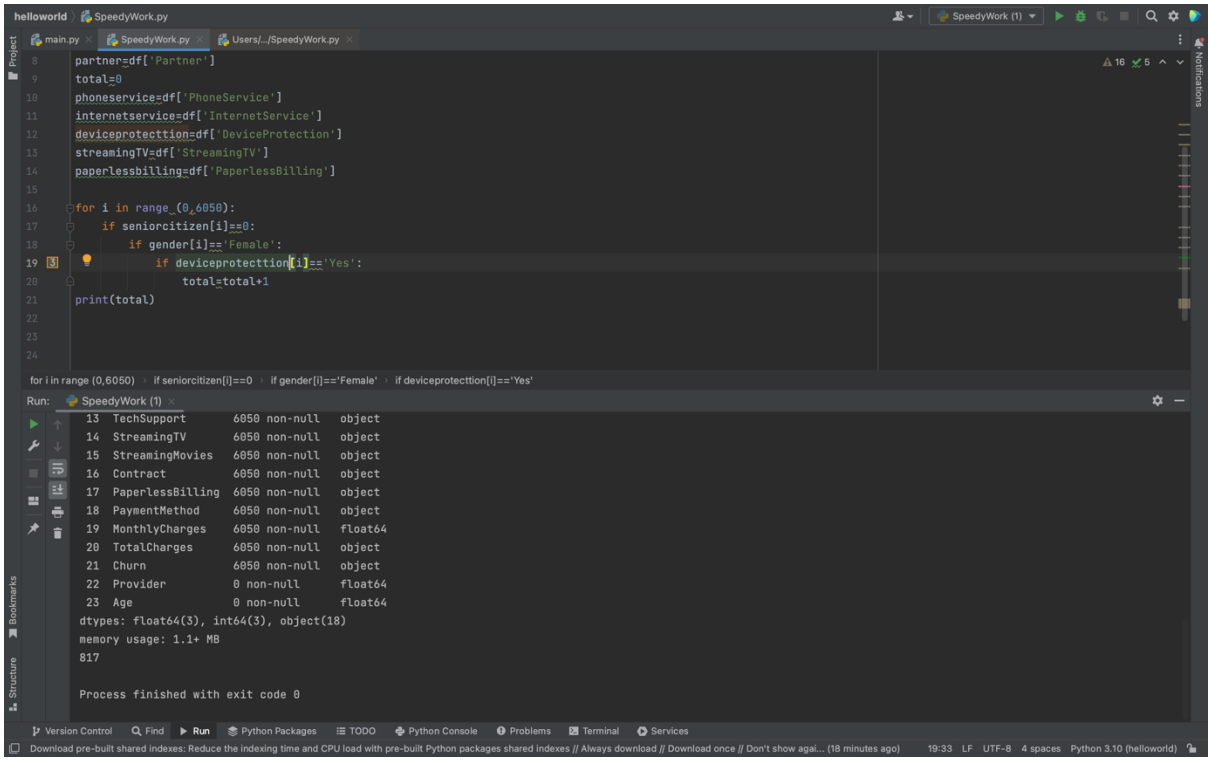
Q2



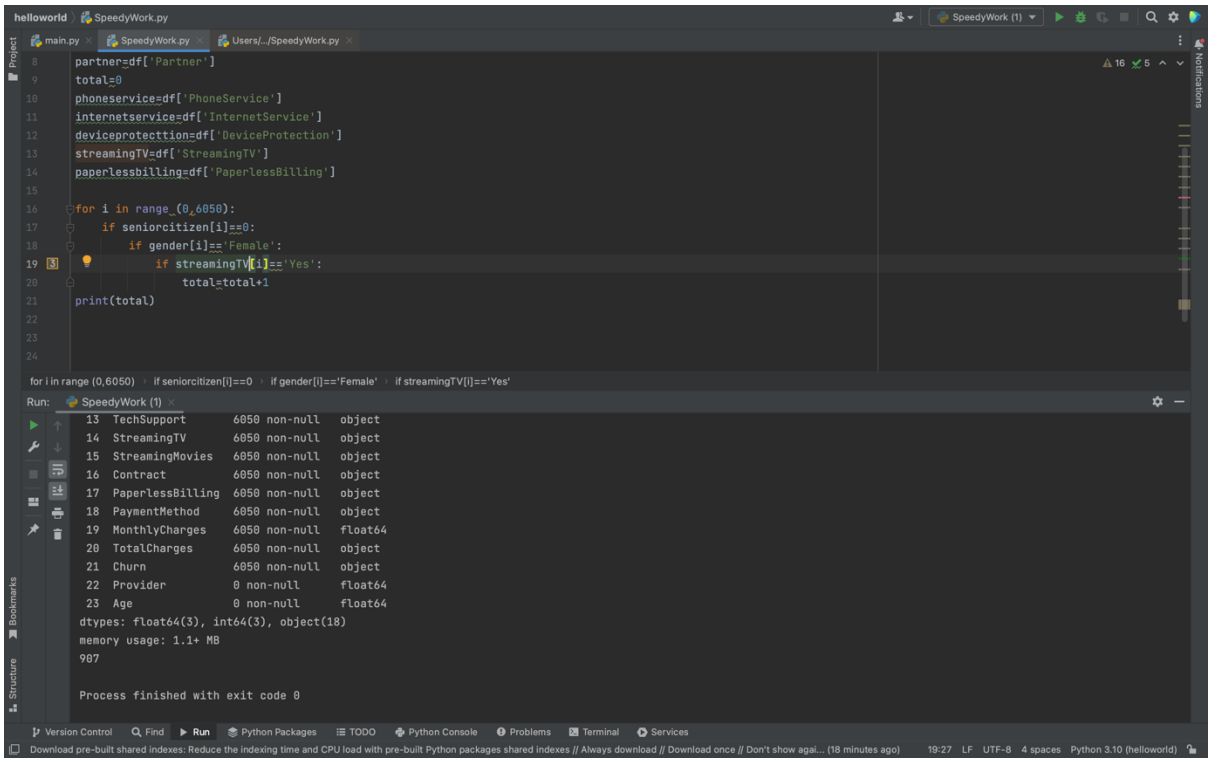
There are 2478 non-Senior Citizen Females that have either DSL or Fiber Optic Internet service.



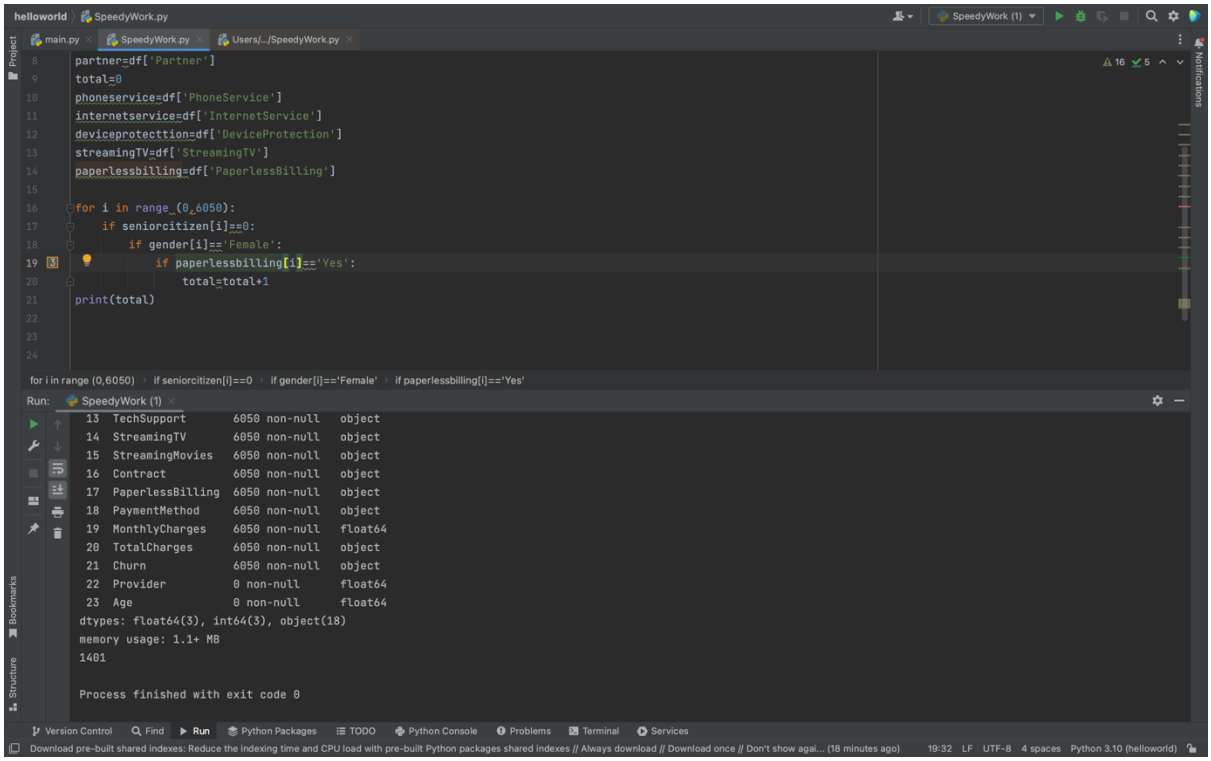
There are 2224 non-Senior Citizen Females that have phone service.



There are 817 non-Senior Citizen Females that have device protection.

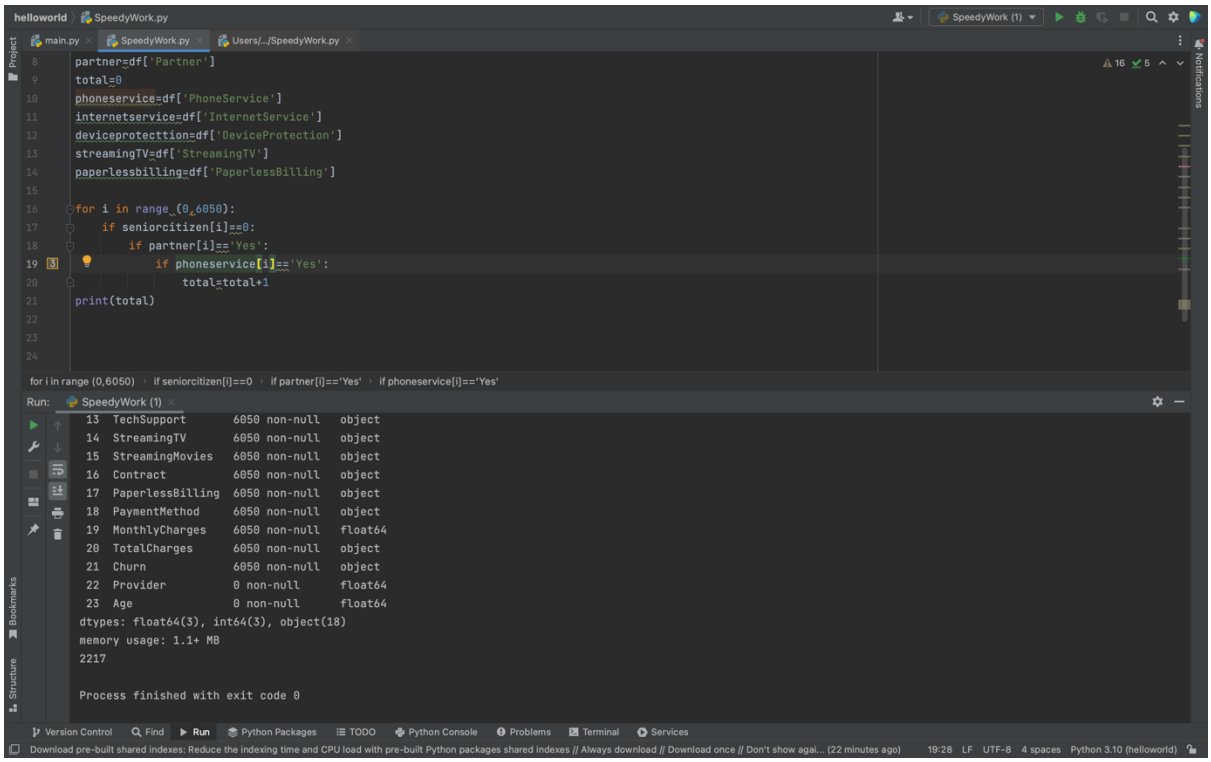


There are 907 non-Senior Citizen Females that have a streaming TV.

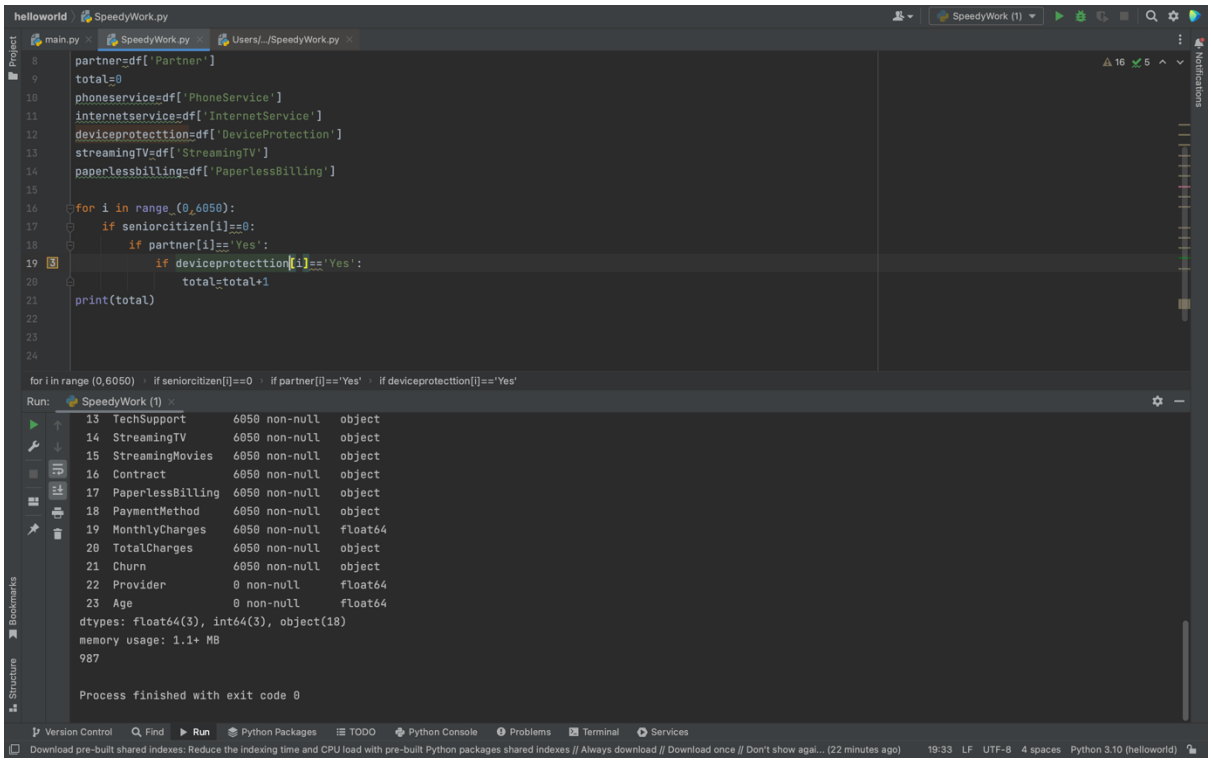


There are 1401 non-Senior Citizen Females that have paperless billing.

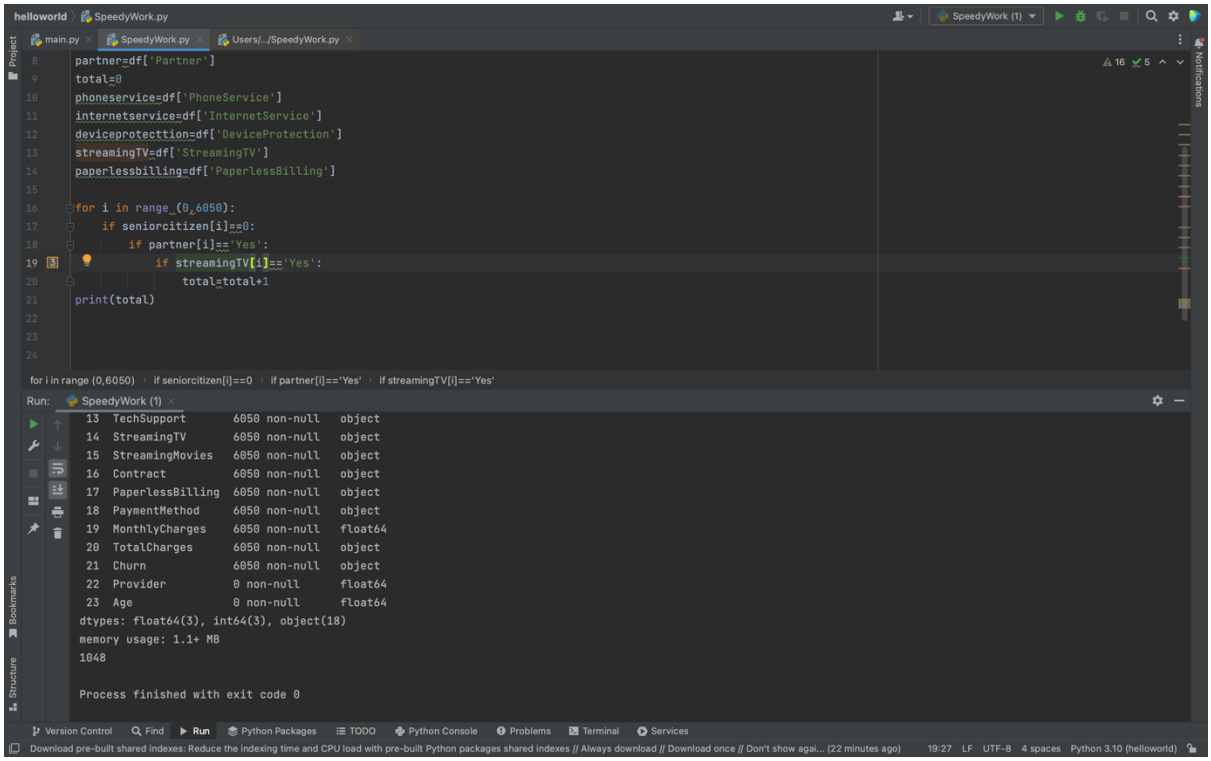
Q3



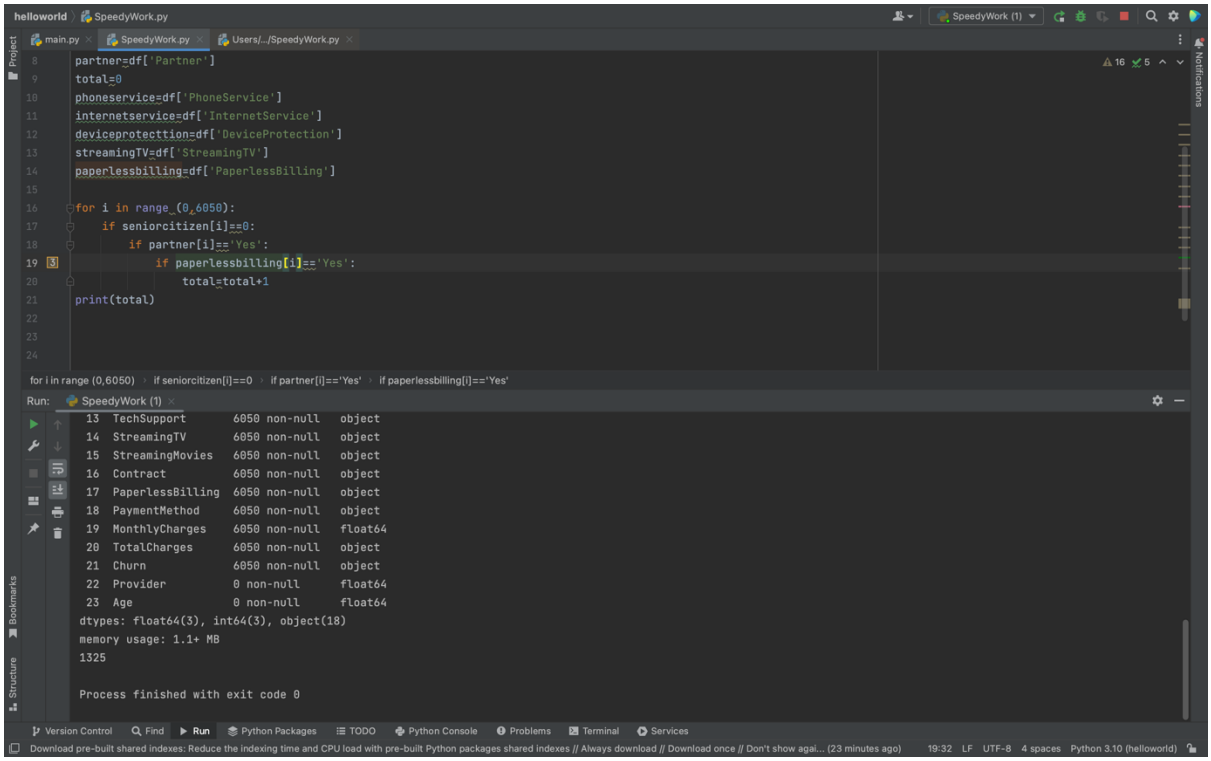
There are 2217 non-Senior Citizen people with partners that have phone service.



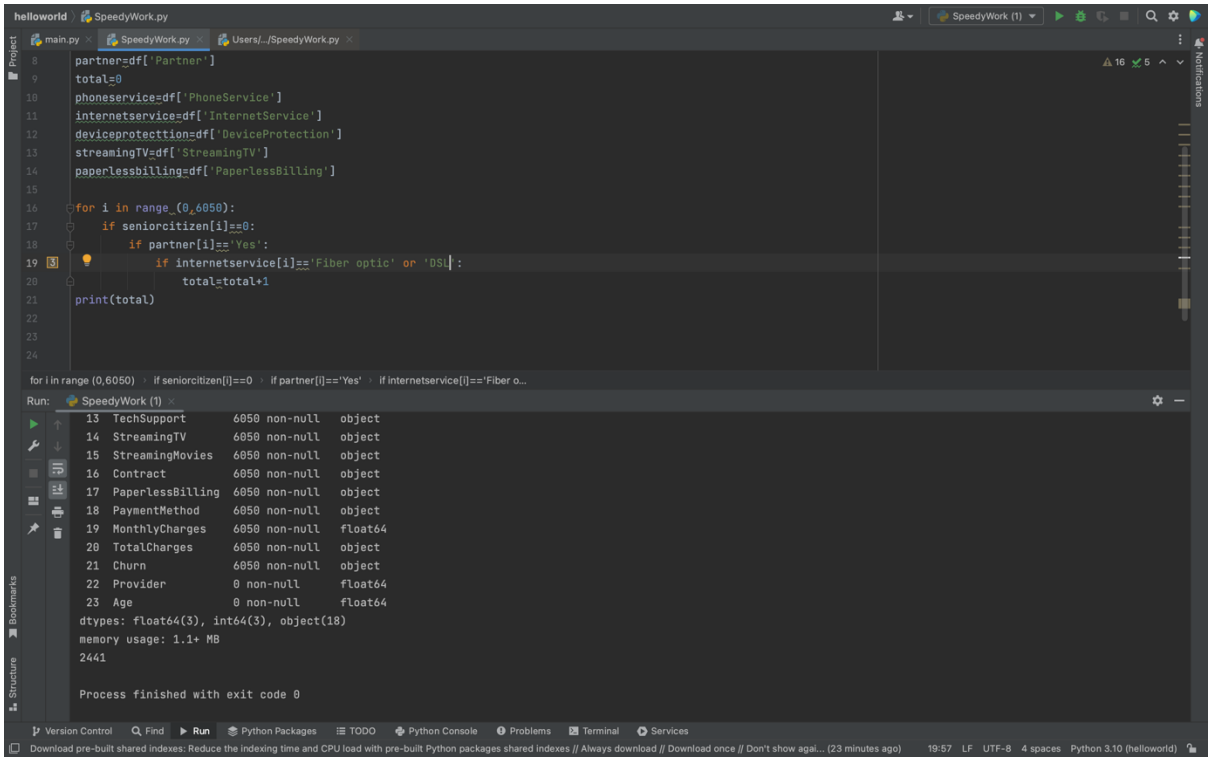
There are 987 non-Senior Citizen people with partners that have device protection.



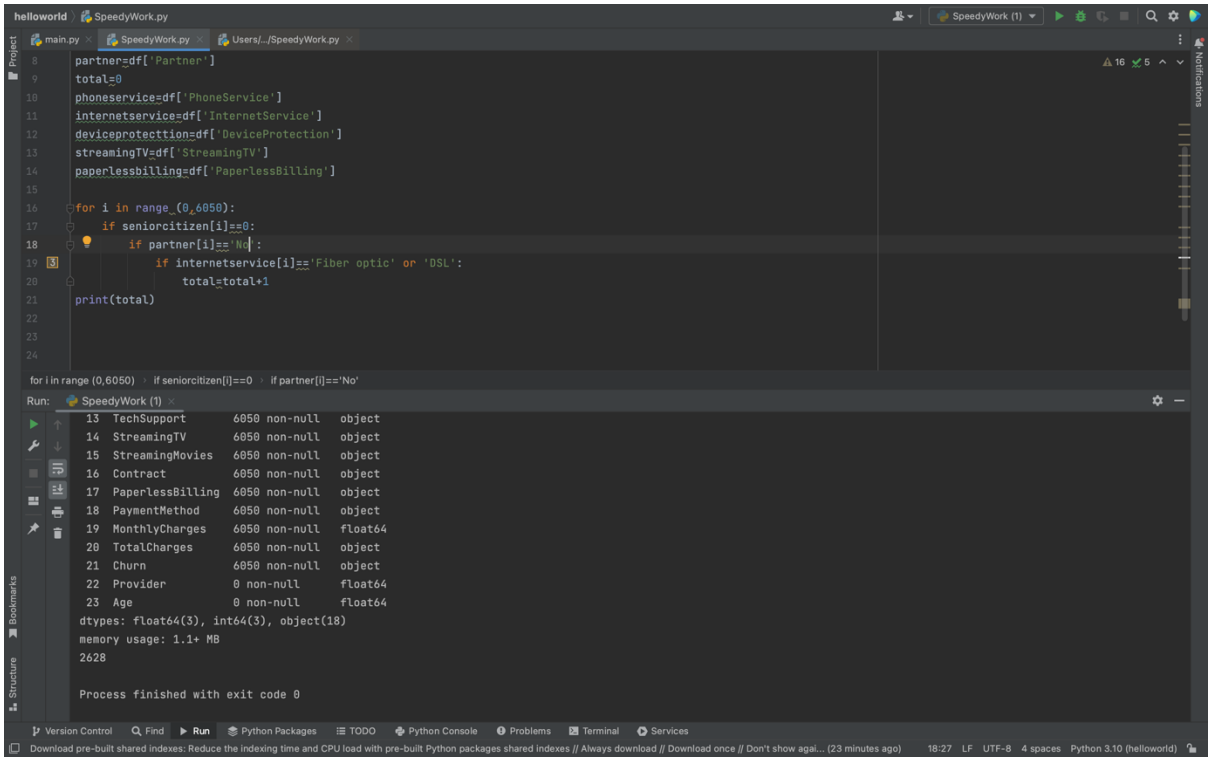
There are 1048 non-Senior Citizen people with partners that have a streaming TV.



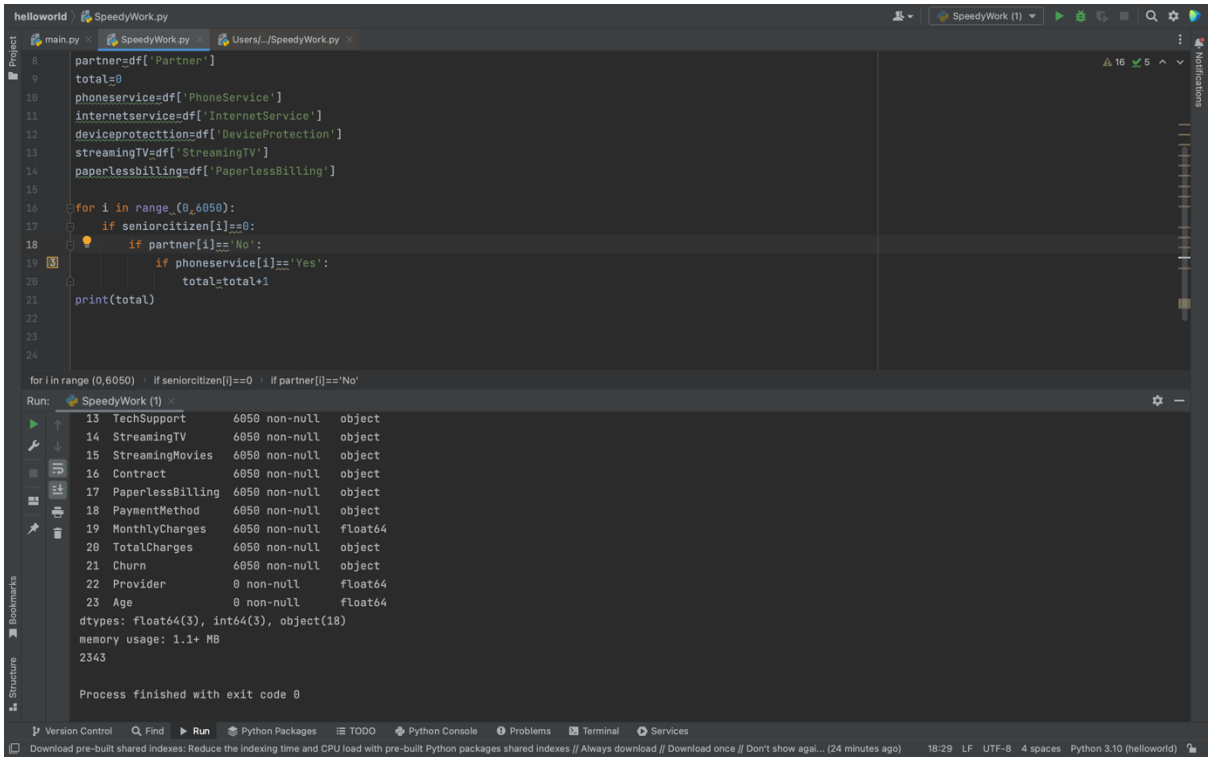
There are 1325 non-Senior Citizen people with partners that have paperless billing.



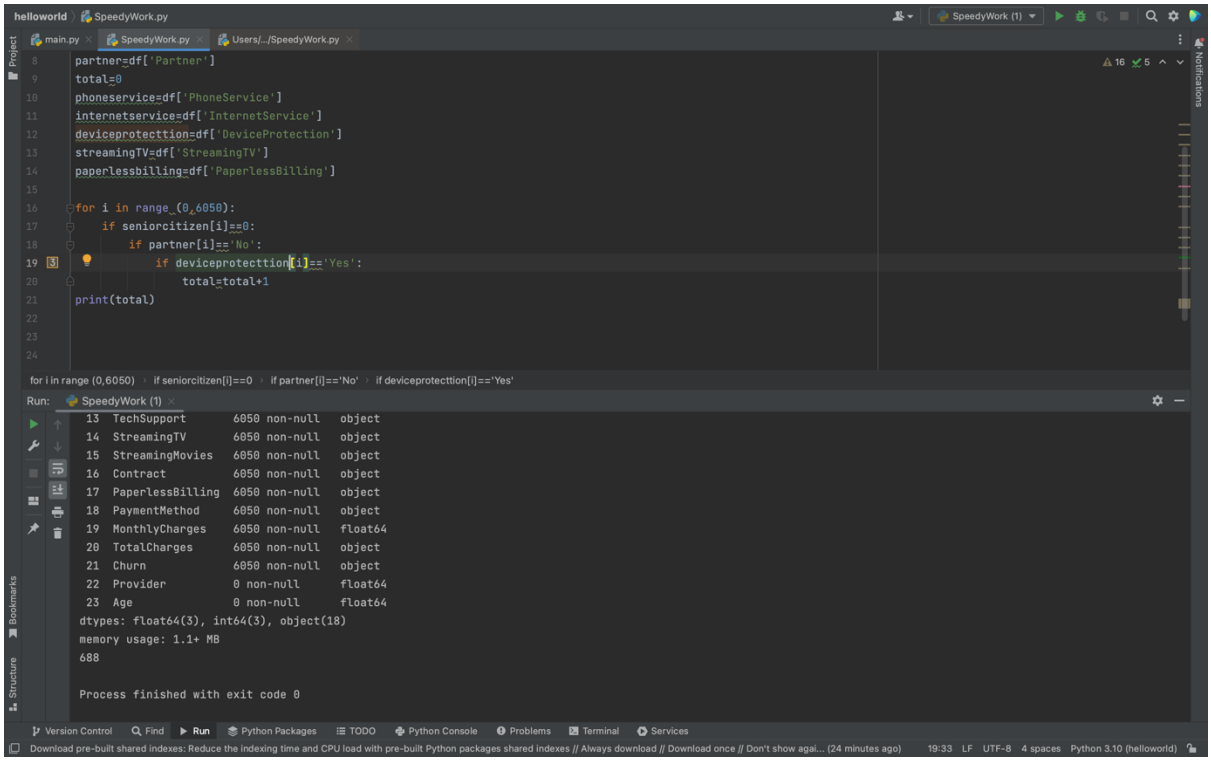
There are 2441 non-Senior Citizen people with partners that have fiber optic or DSL as their internet service.



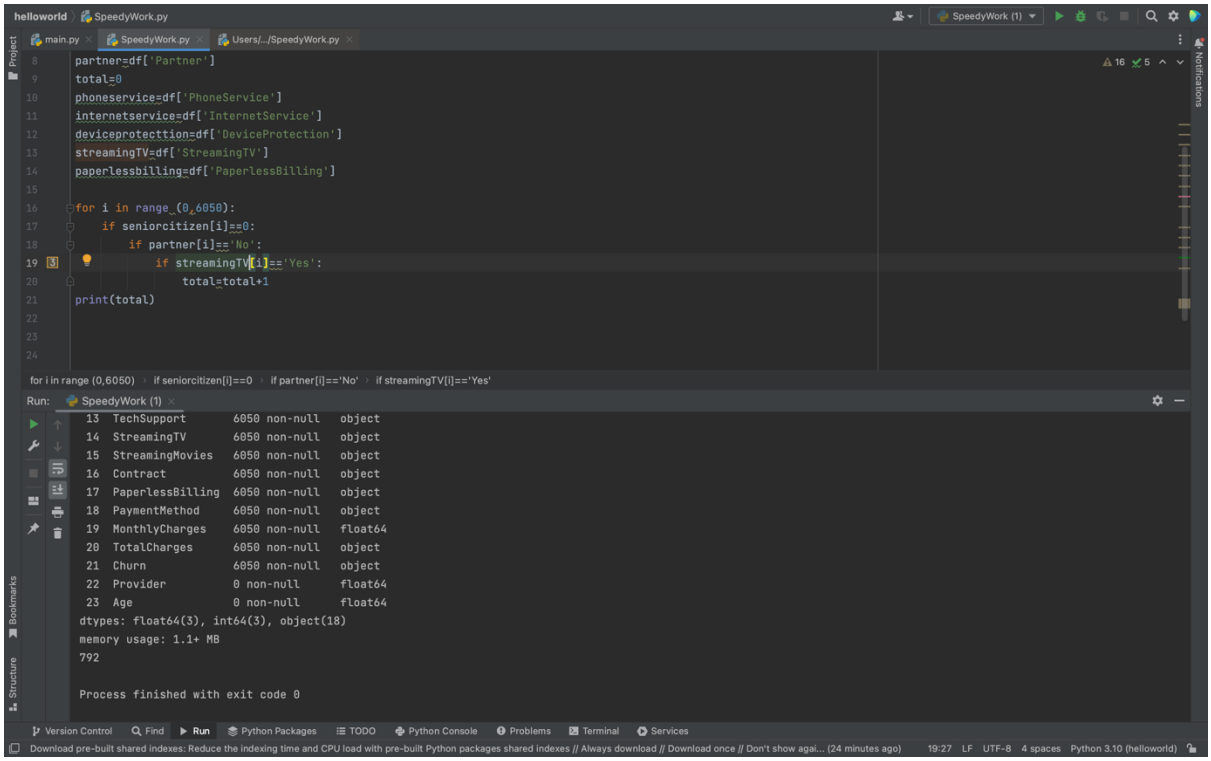
There are 2628 non-Senior Citizen people without partners that have fiber optic or DSL.



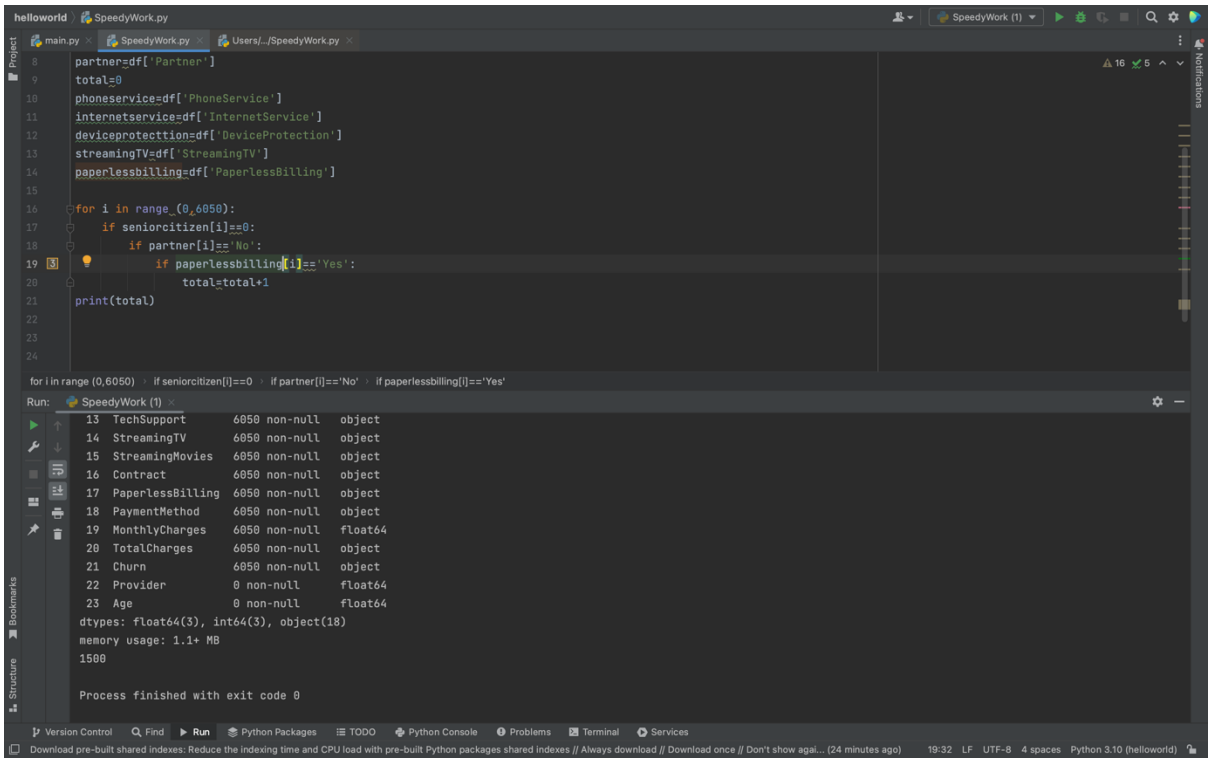
There are 2343 non-Senior Citizen people without partners that have phone service



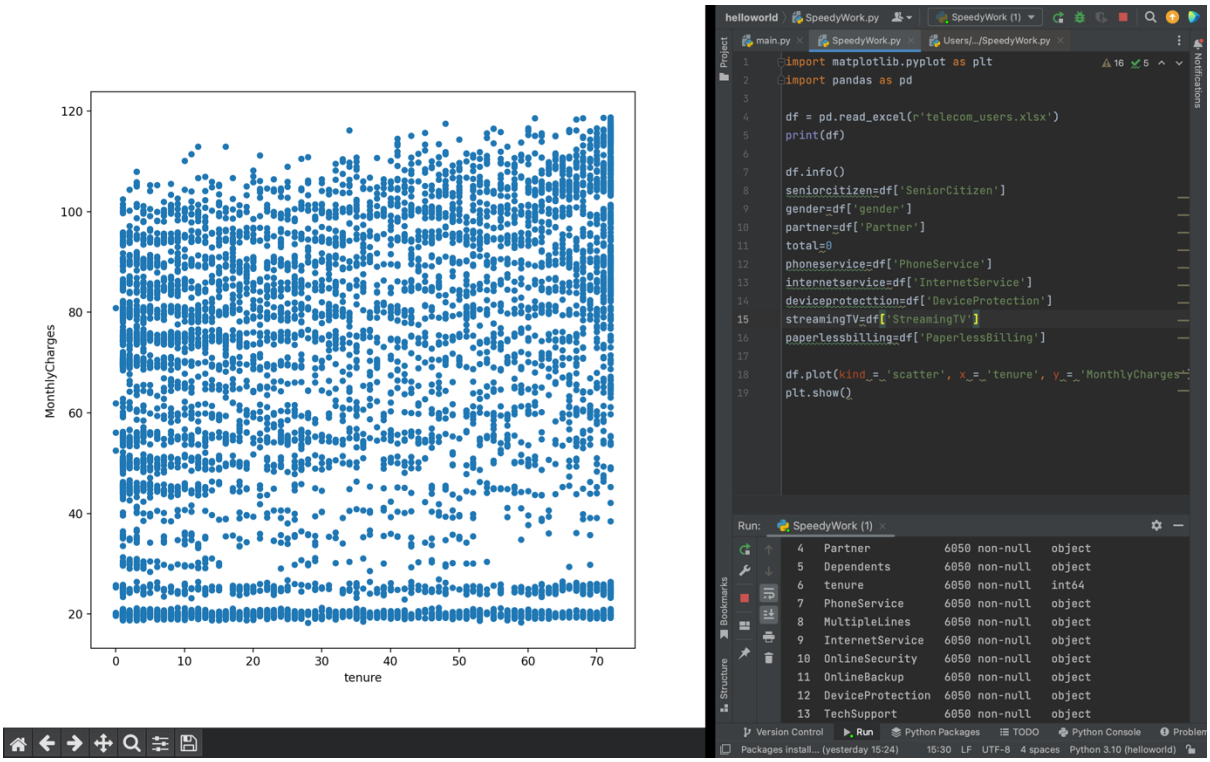
There are 688 non-Senior Citizen people without partners that have device protection.



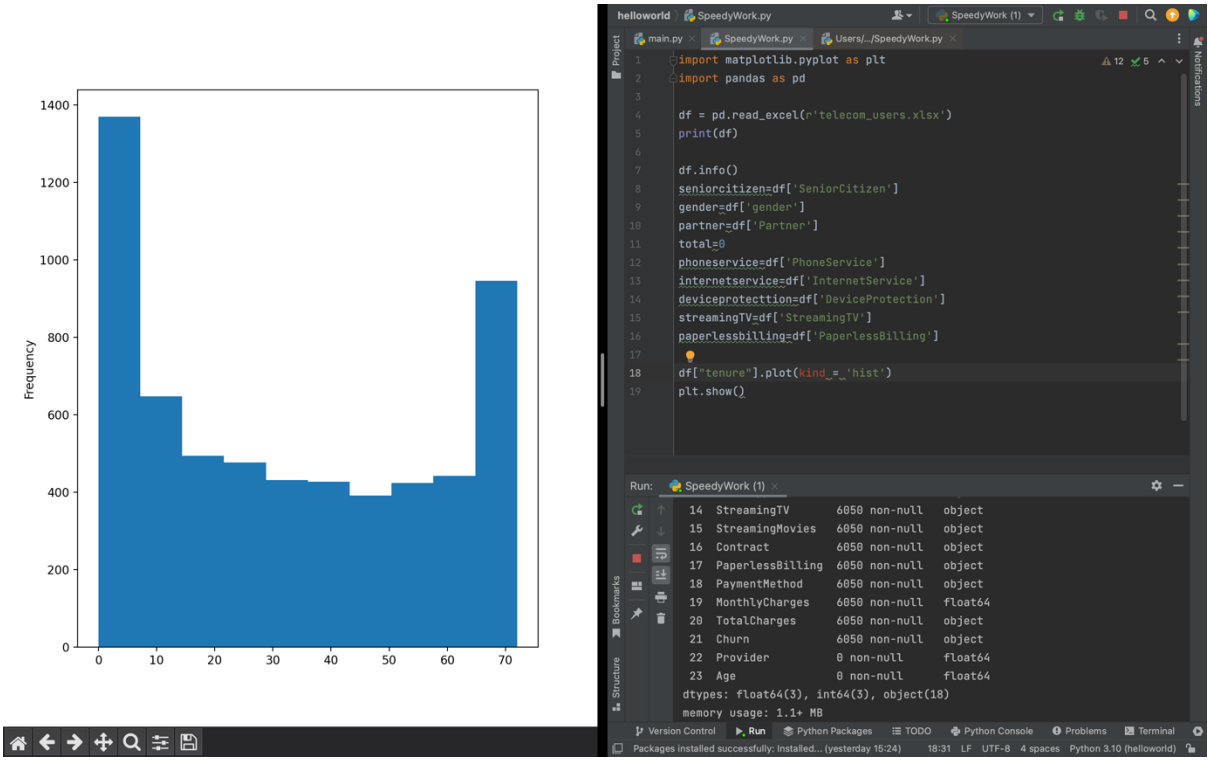
There are 792 non-Senior Citizen people without partners that have a streaming TV.



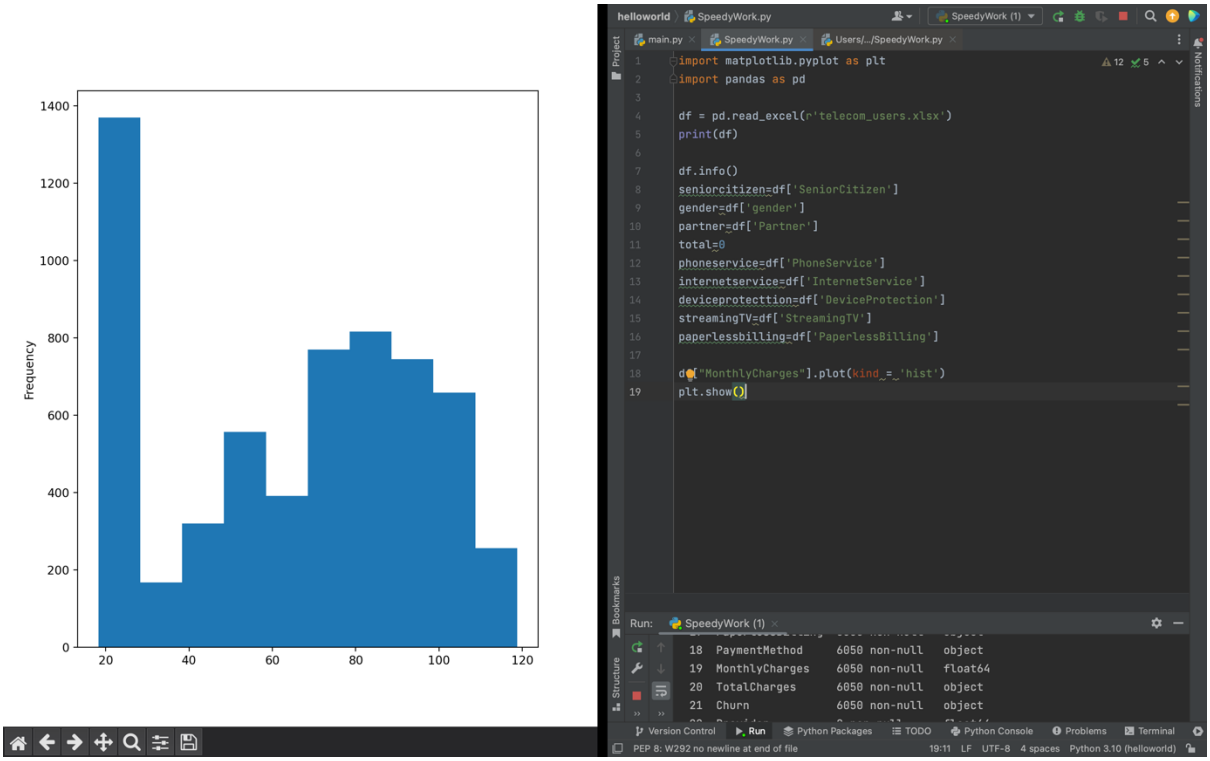
There are 1500 non-Senior Citizen people without partners that have paperless billing.



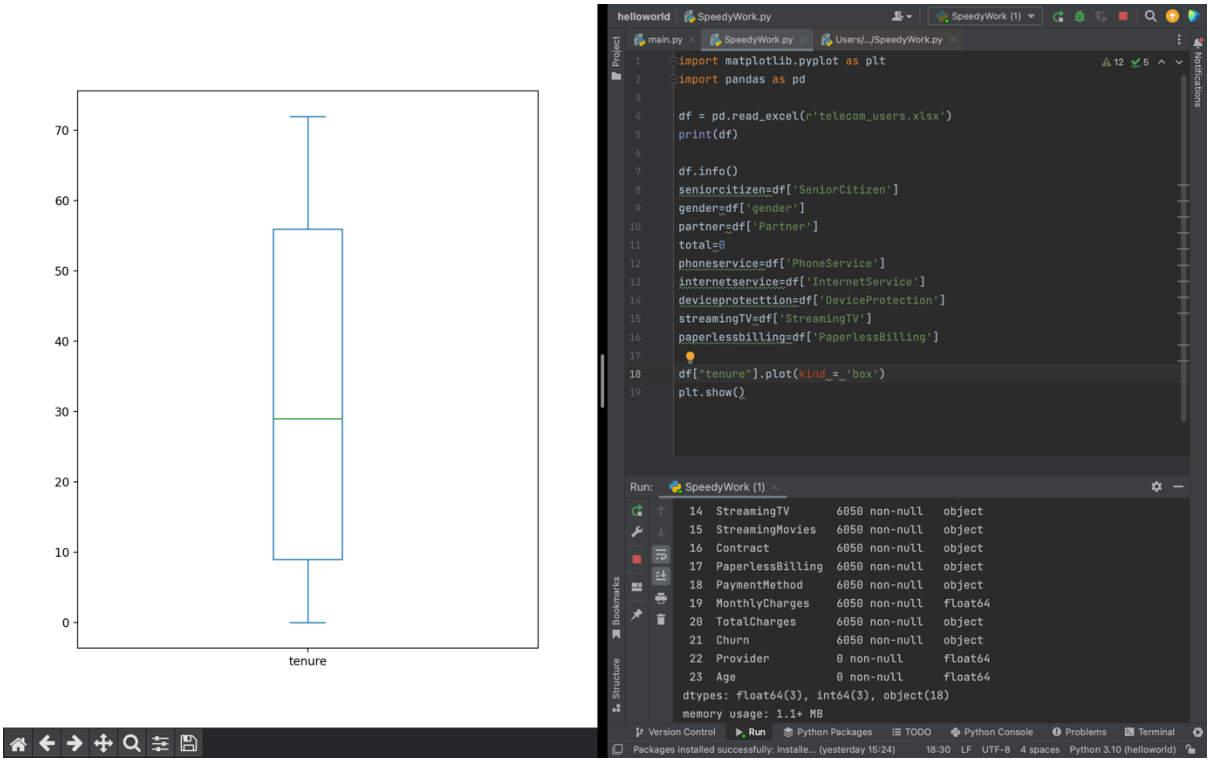
Scatter graph created using python. The graph looks at the results of people's tenure compared to their monthly charge with no clear correlation between the two.



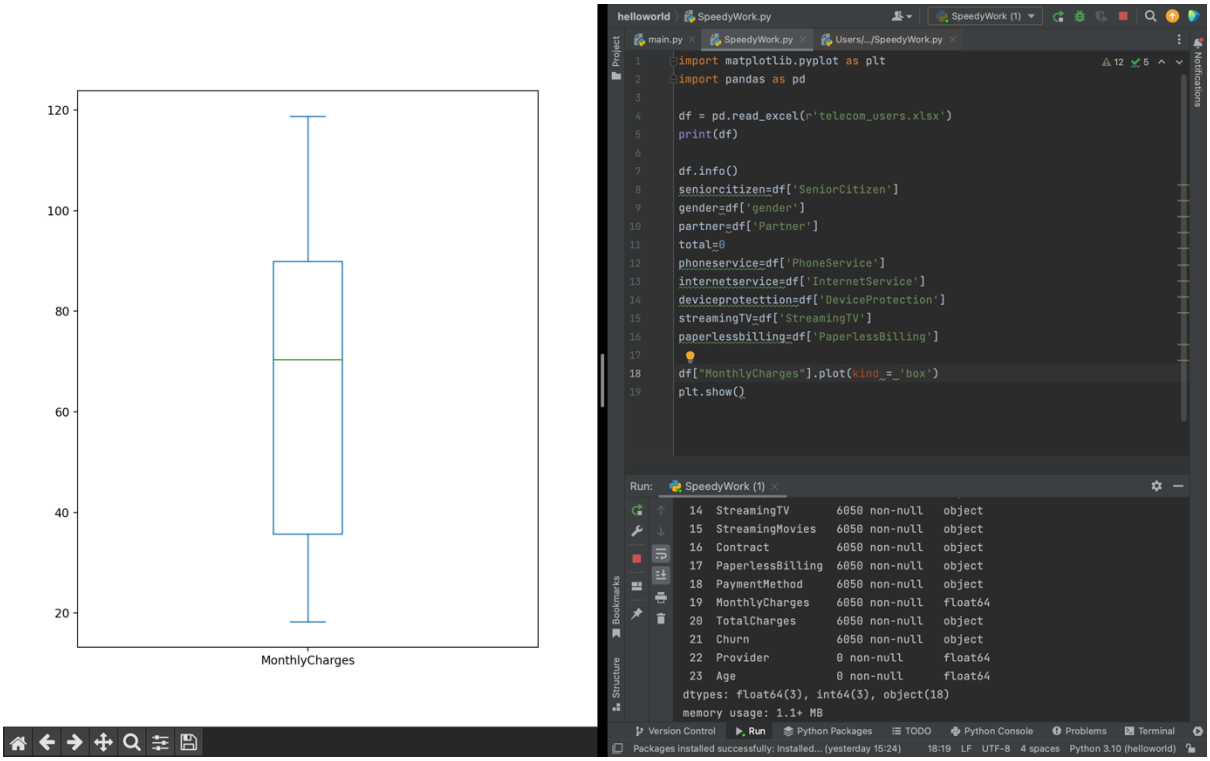
Histogram created used python to show the frequency of different tenures amongst the people in the data.



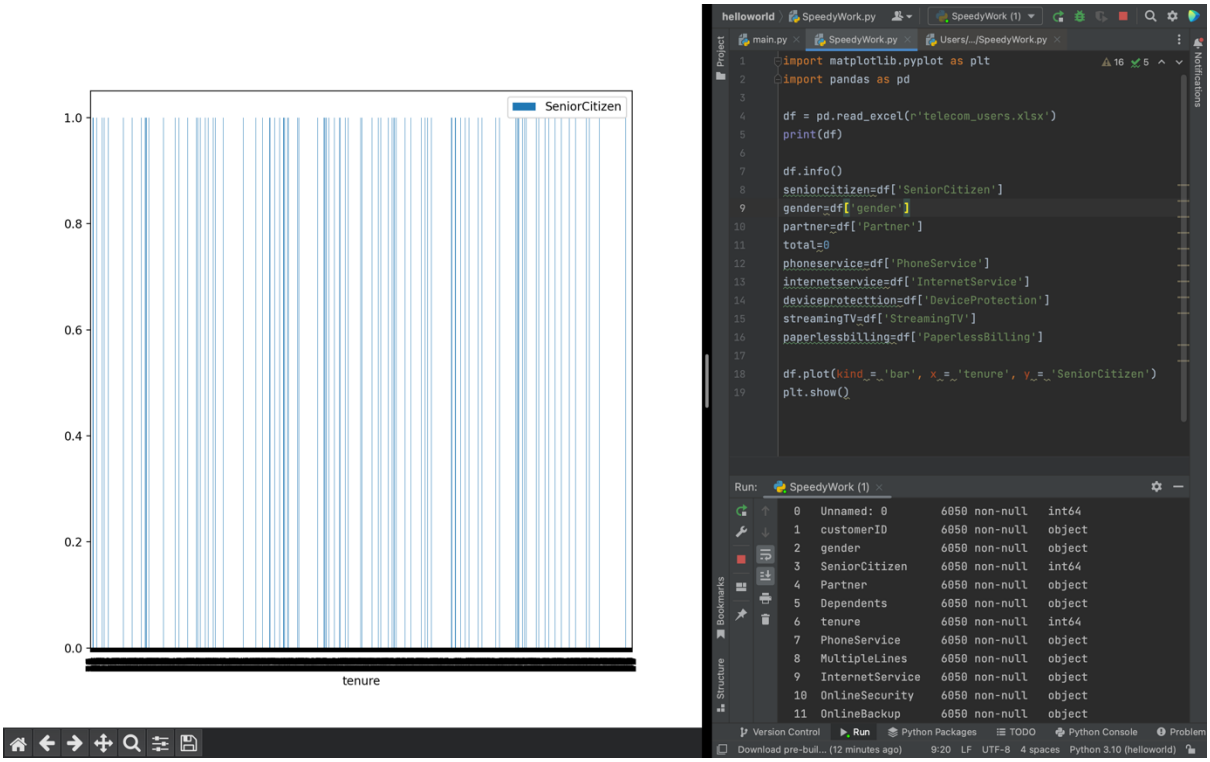
Histogram created used python to show the frequency of different monthly charges amongst the people in the data.



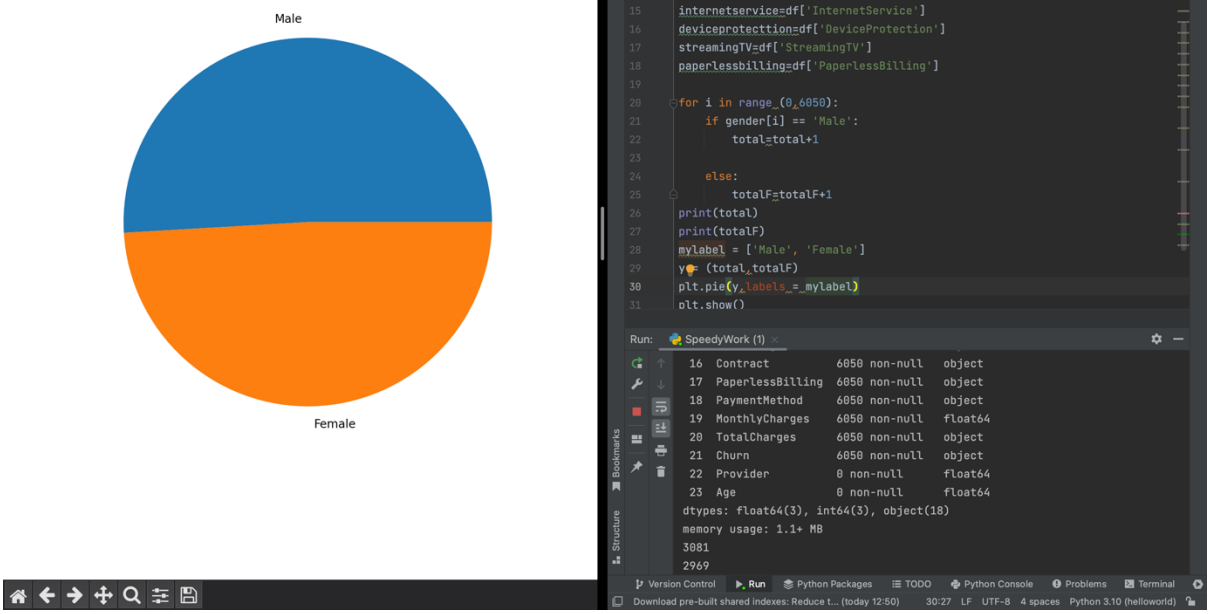
Boxplot created used python to show the data on peoples tenure in the data. The median tenure appears to be slightly under 30, while the longest tenure is over 70.



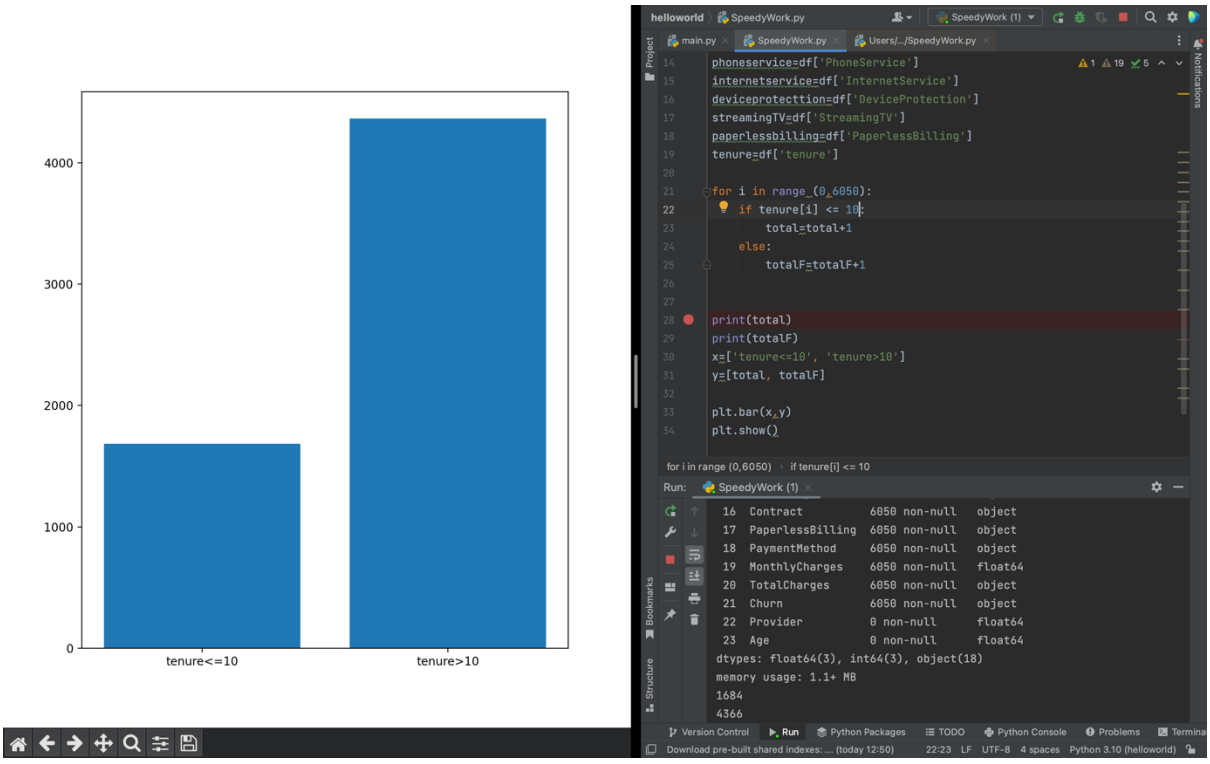
Boxplot created used python to show the data on peoples monthly charges in the data. At a minimum, people in the data are paying around 18. However, majority of the people in the data pay between 38 to 88.



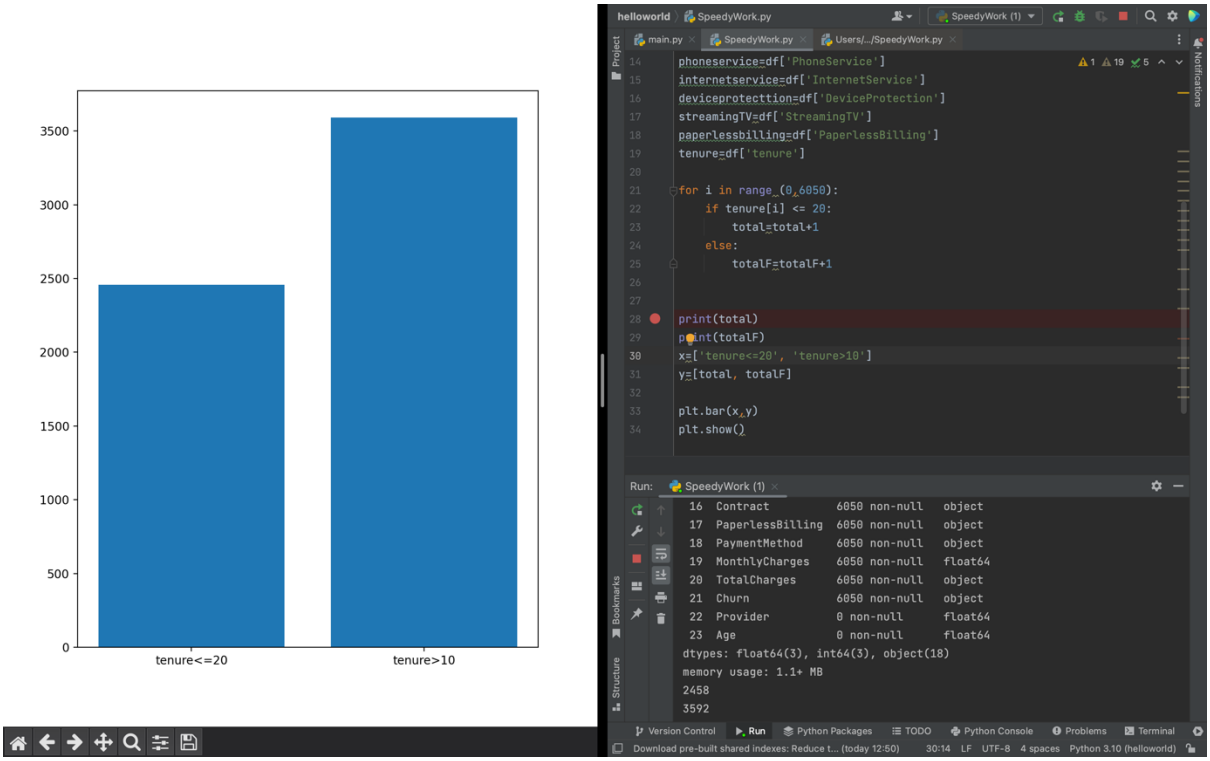
Scatter graph plotting the tenure of senior citizens.



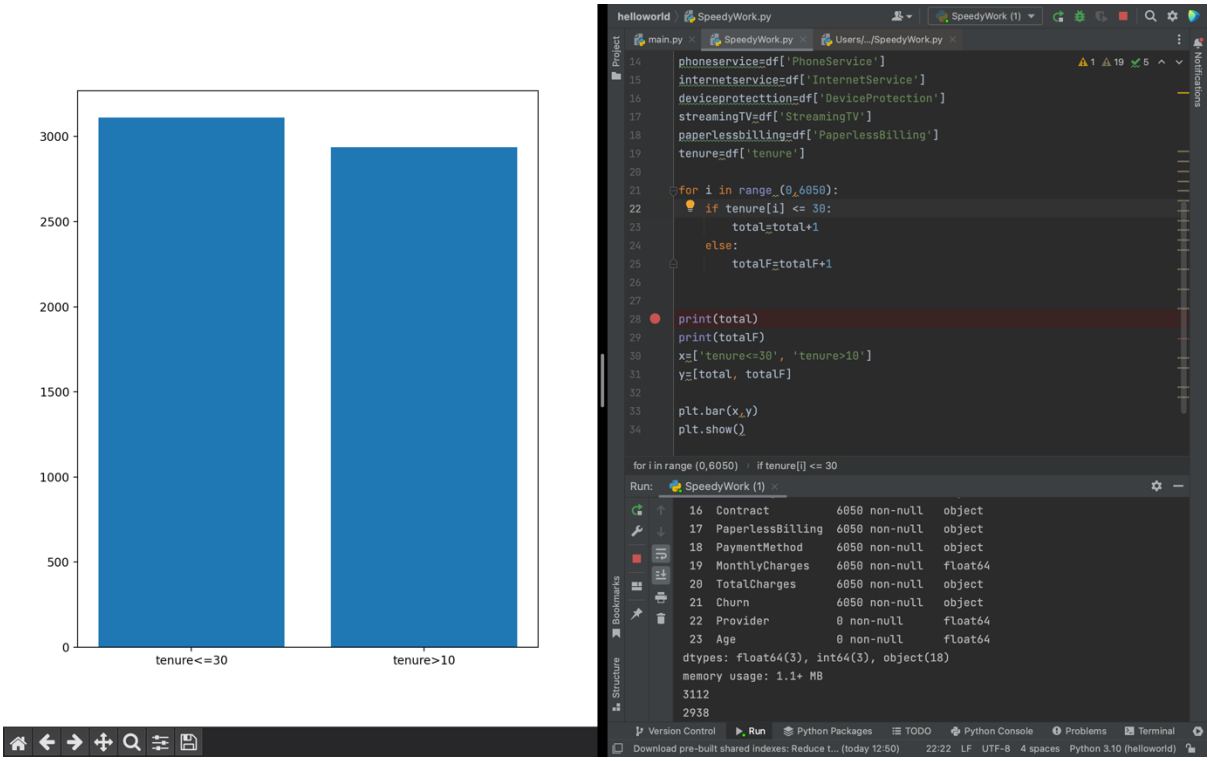
Pie chart depicting the share of males and females in the dataset.



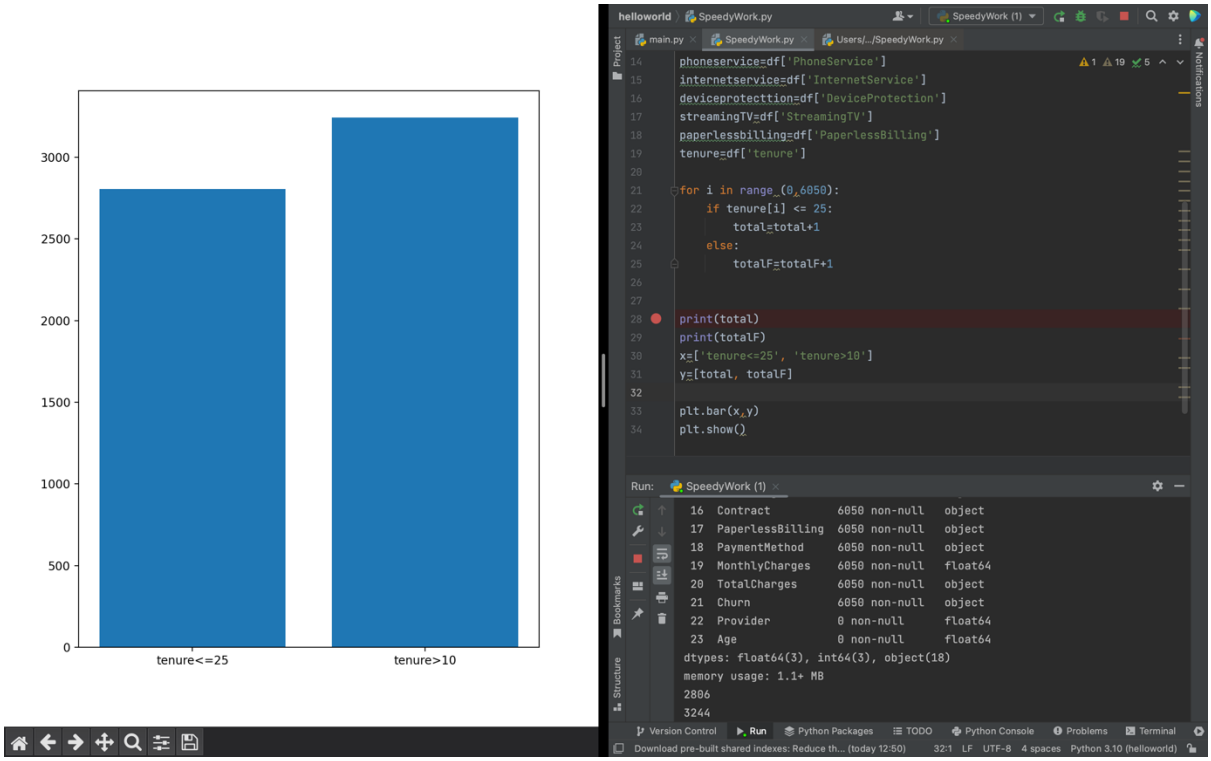
Bar chart comparing the number of people in the data set who had a tenure shorter than 10 and everyone else in the data.



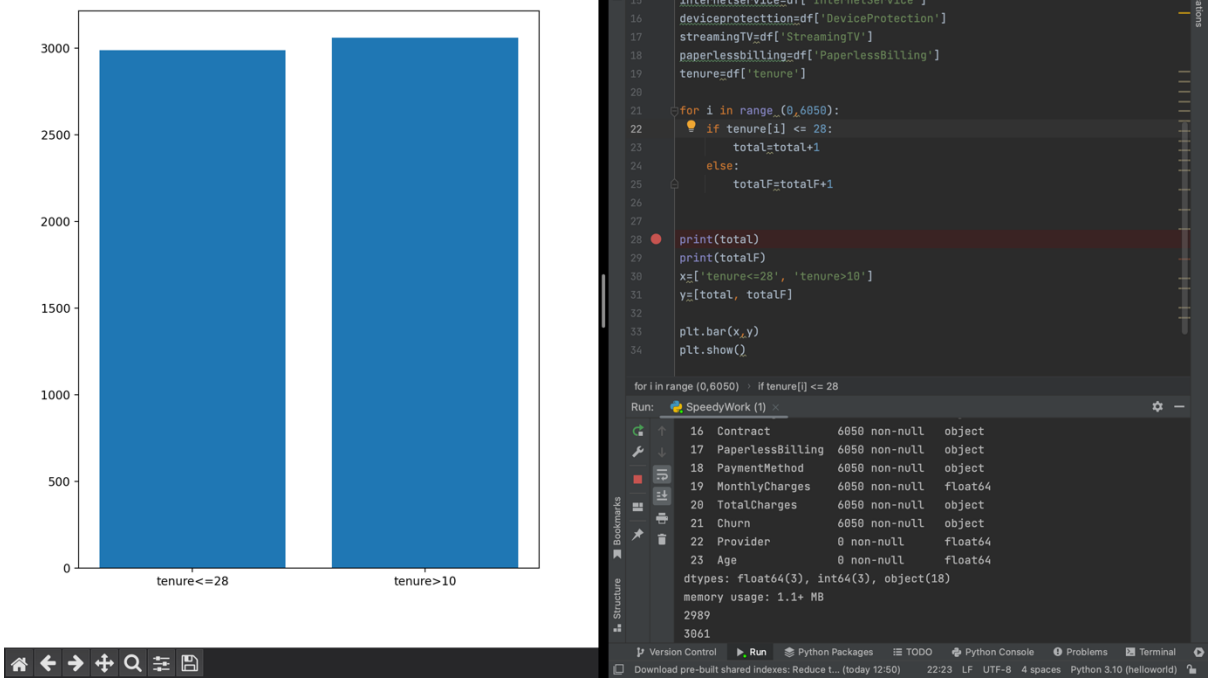
Bar chart comparing the number of people in the data set who had a tenure shorter than 20 and everyone else in the data.



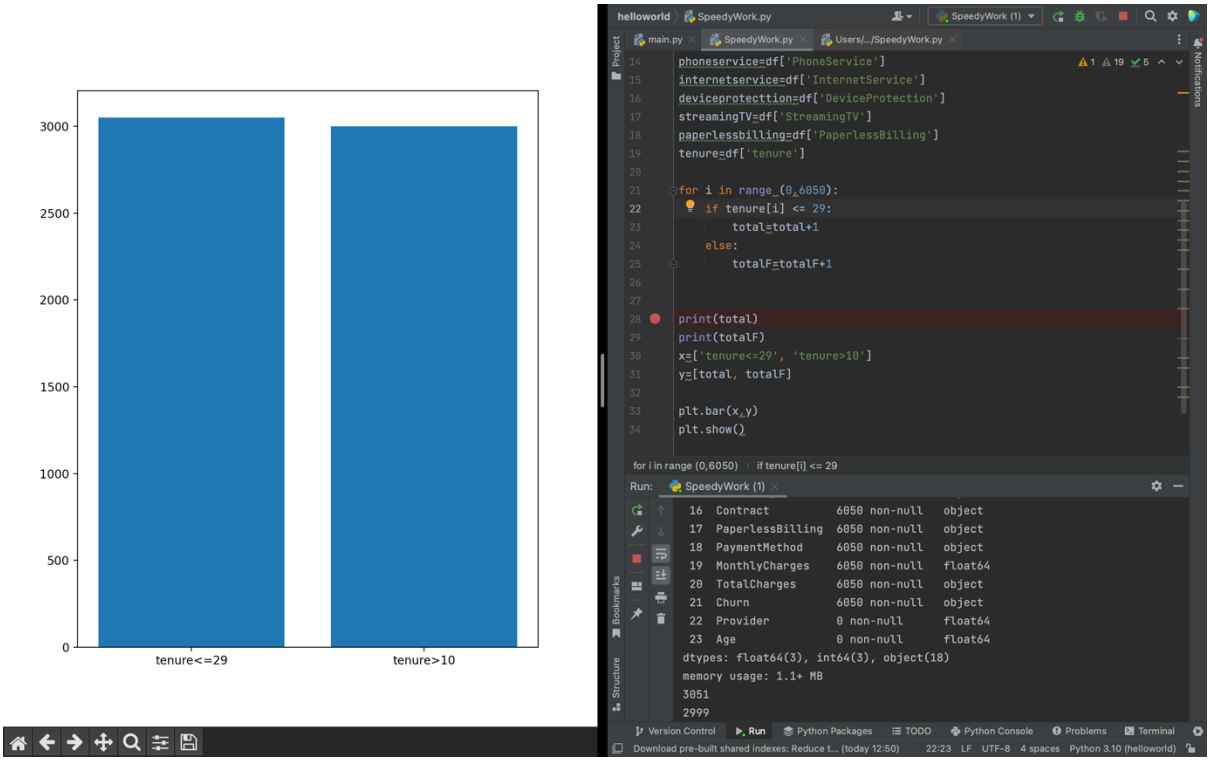
Bar chart comparing the number of people in the data set who had a tenure shorter than 30 and everyone else in the data. This shows that the majority of people in this data set have a tenure of less than 30, compared to the population of higher tenures.



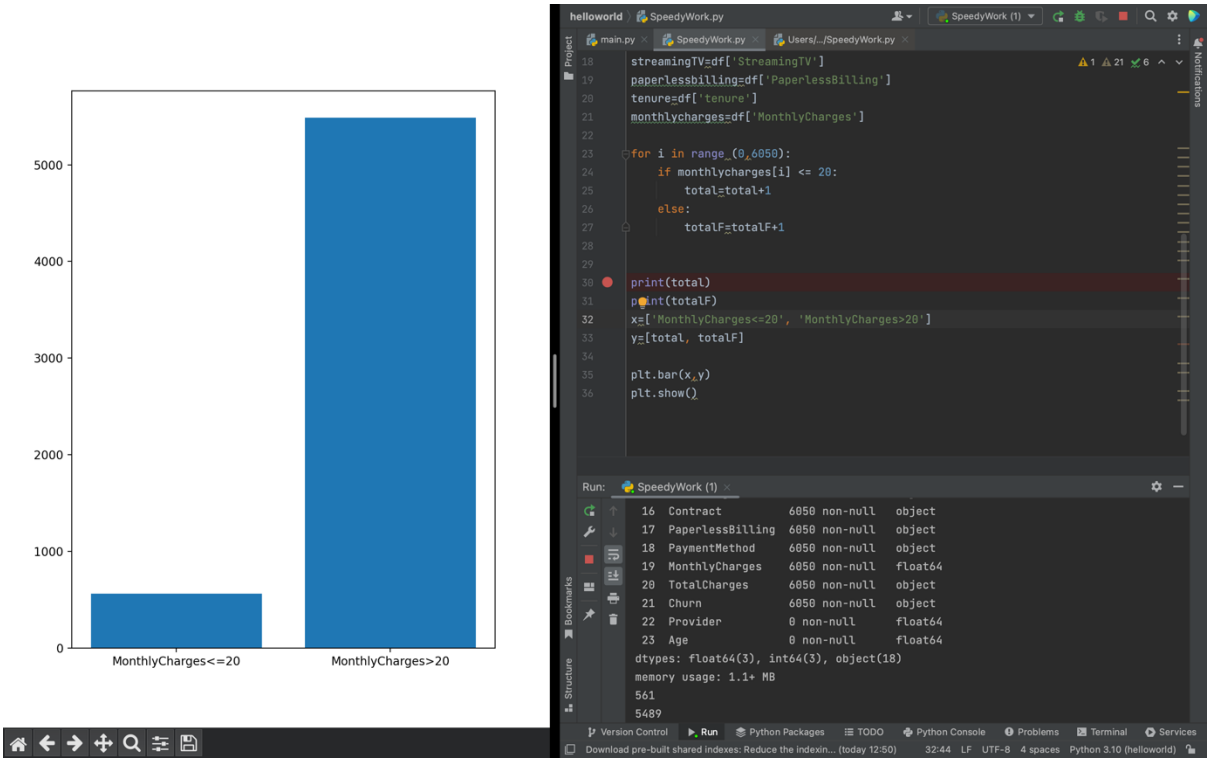
Bar chart comparing the number of people in the data set who had a tenure shorter than 25 and everyone else in the data. This was done to more accurately find the point at which the majority of people lay in.



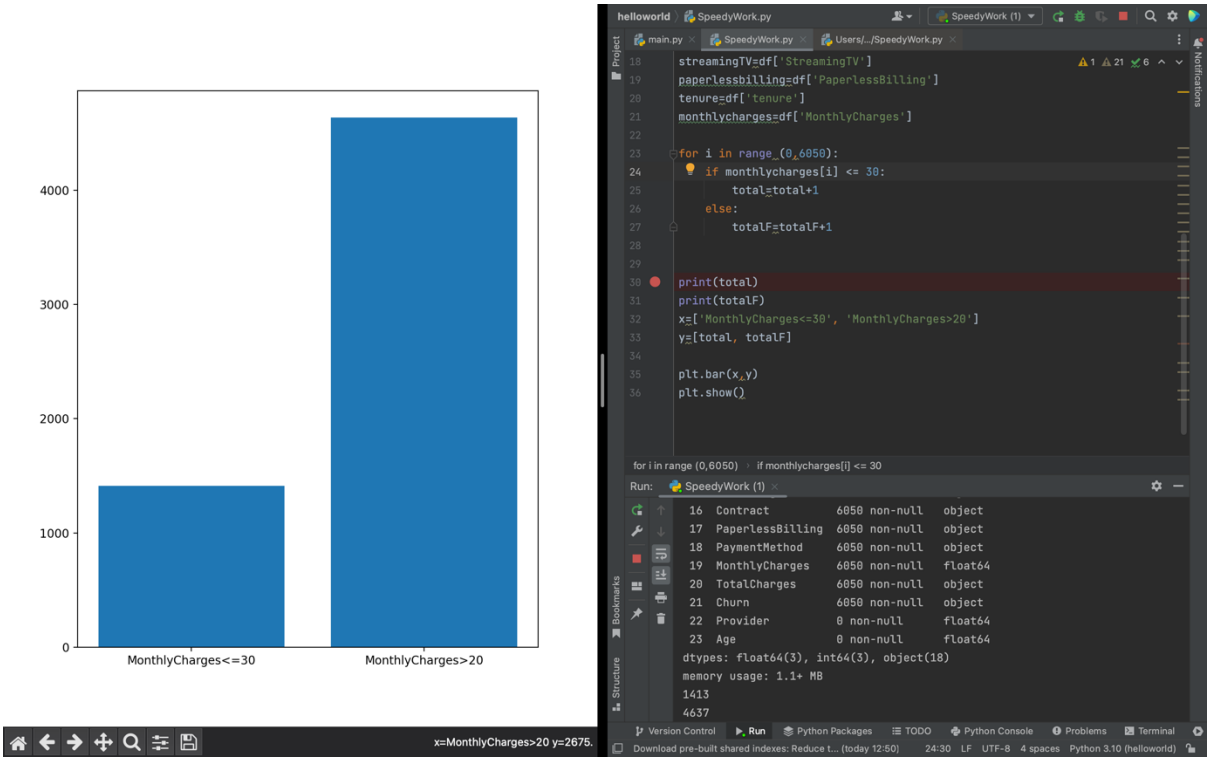
Bar chart comparing the number of people in the data set who had a tenure shorter than 28 and everyone else in the data. This was done to more accurately find the point at which the majority of people lay in. Cut off point is 29.



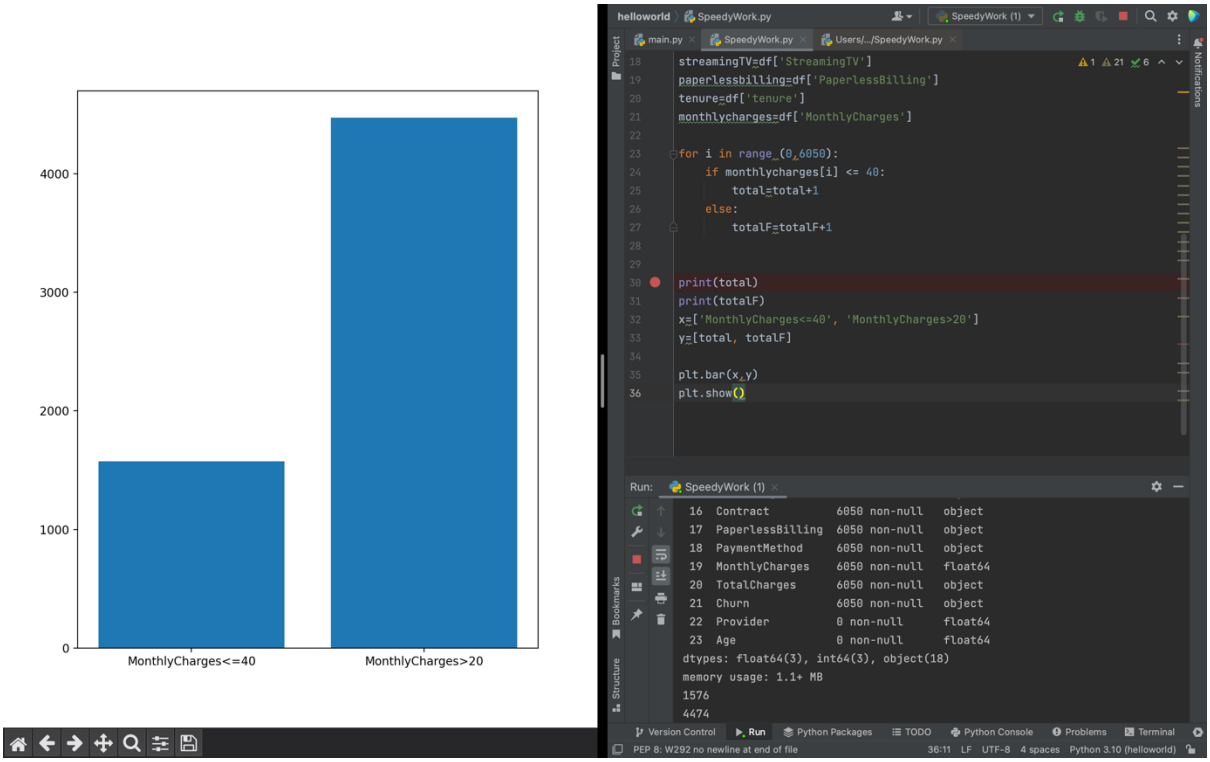
Bar chart comparing the number of people in the data set who had a tenure shorter than 28 and everyone else in the data. This was done to more accurately find the point at which the majority of people lay in. Cut off point is 29.



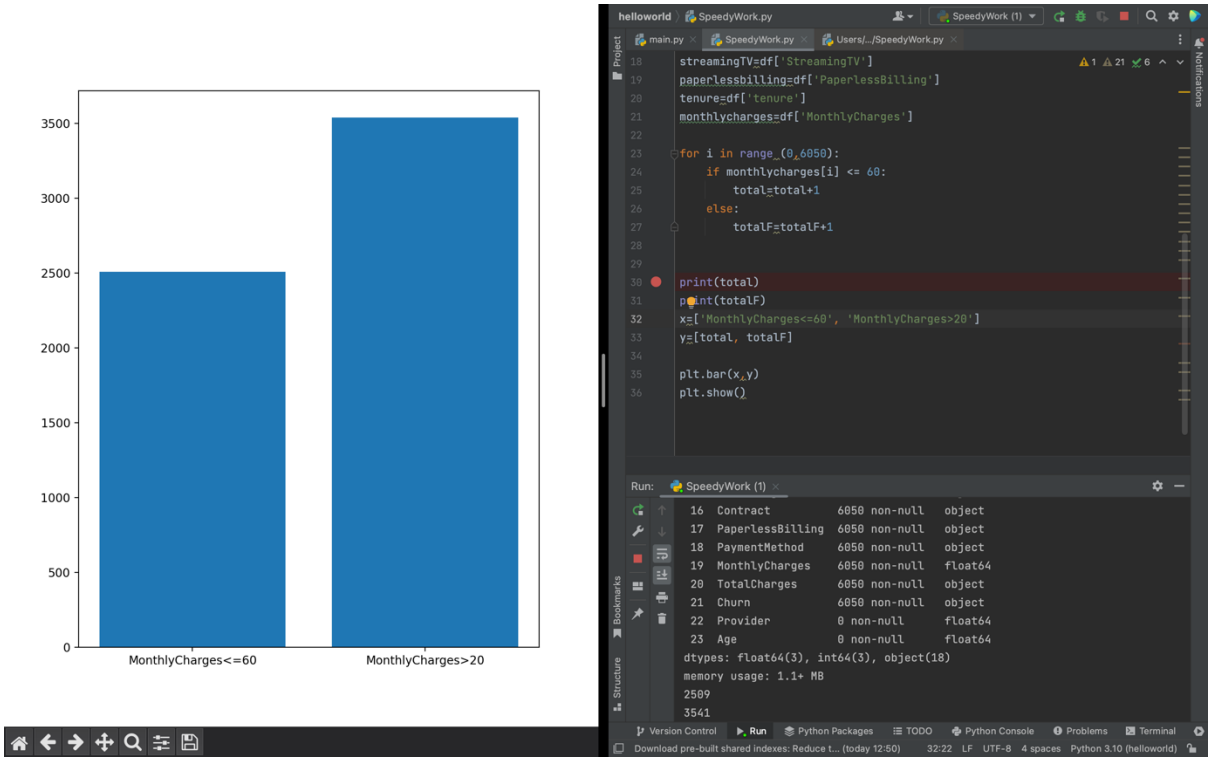
Bar chart comparing the number of people in the data set who had a monthly charge of less than 20 and everyone else in the data.



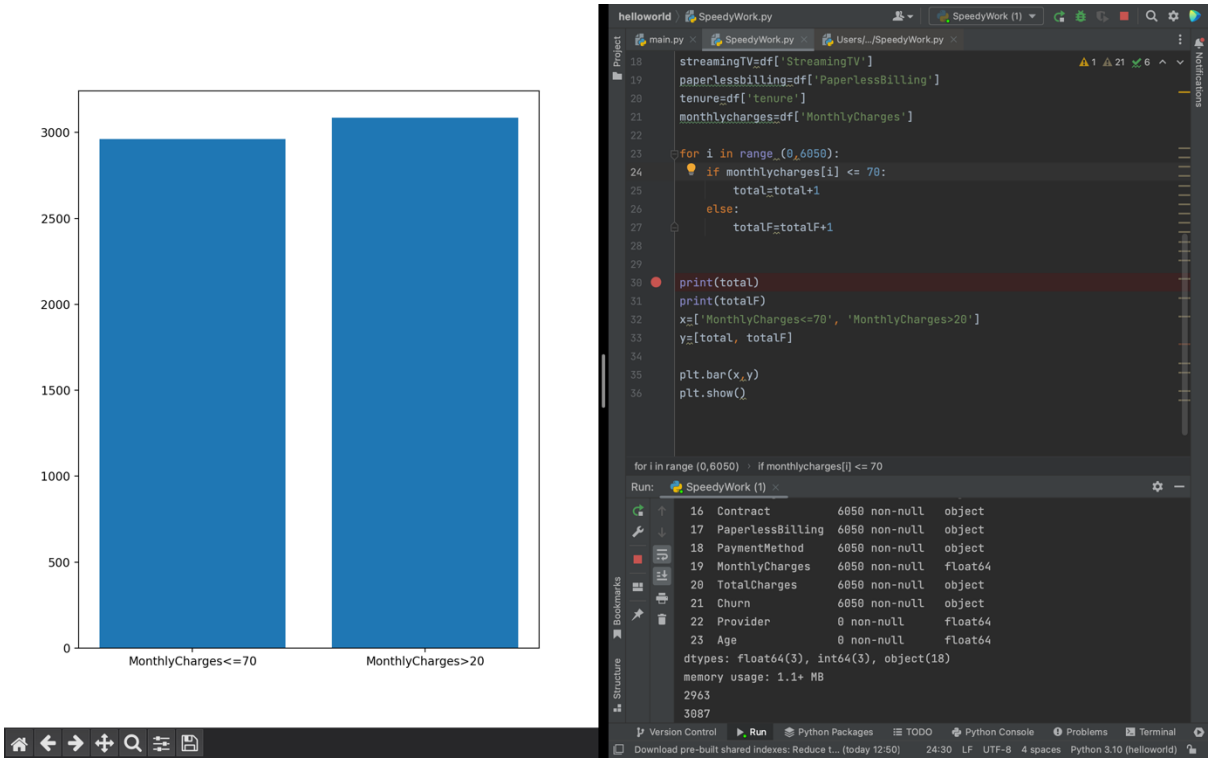
Bar chart comparing the number of people in the data set who had a monthly charge of less than 30 and everyone else in the data.



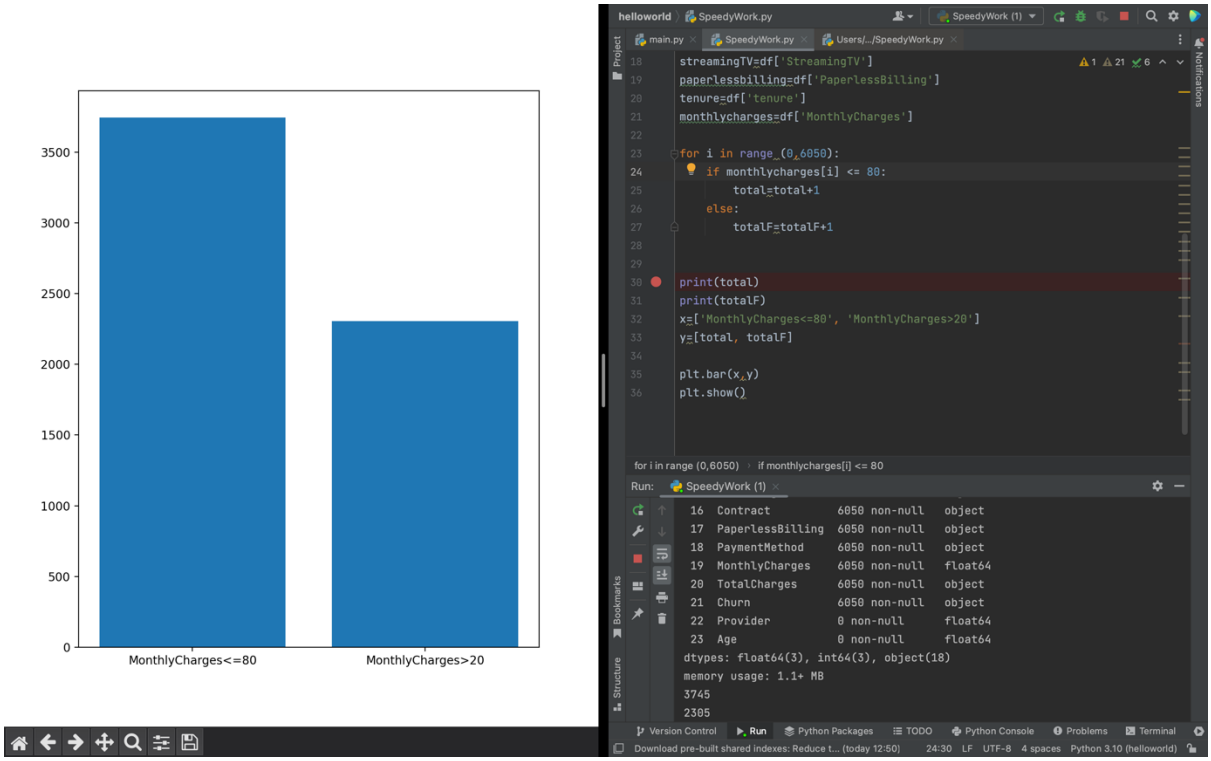
Bar chart comparing the number of people in the data set who had a monthly charge of less than 40 and everyone else in the data.



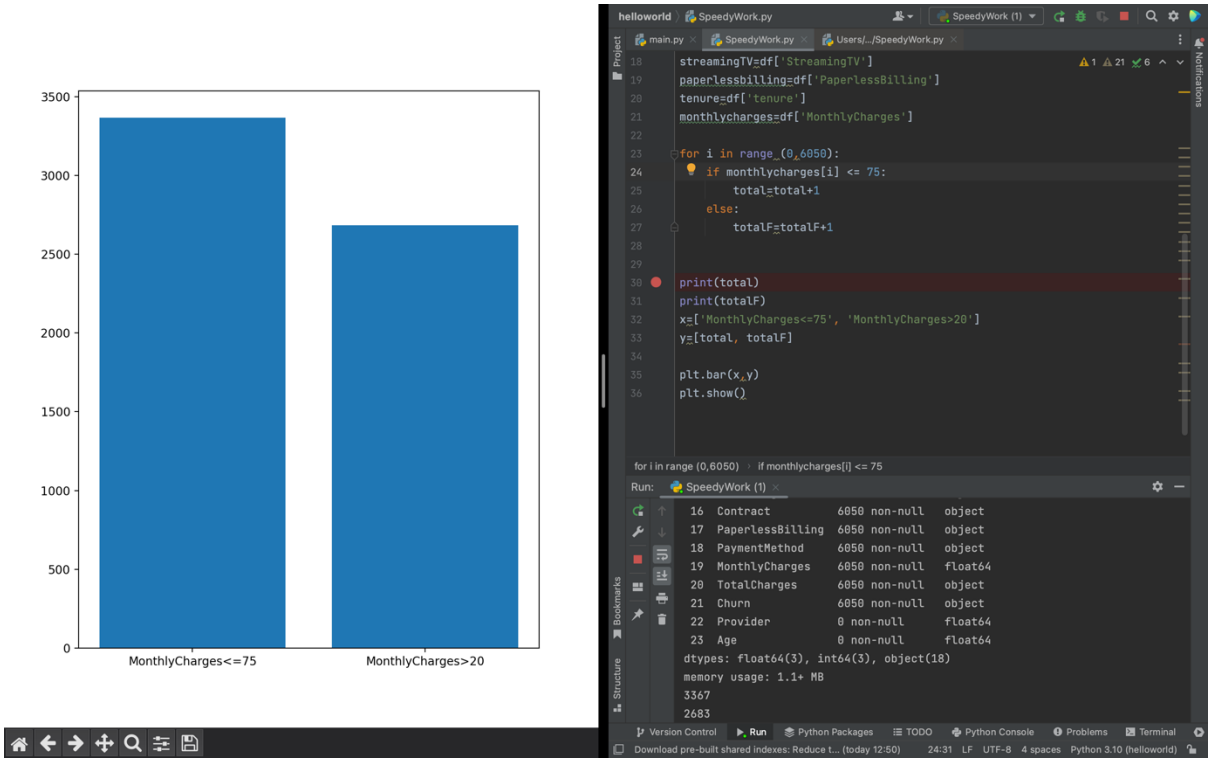
Bar chart comparing the number of people in the data set who had a monthly charge of less than 60 and everyone else in the data.



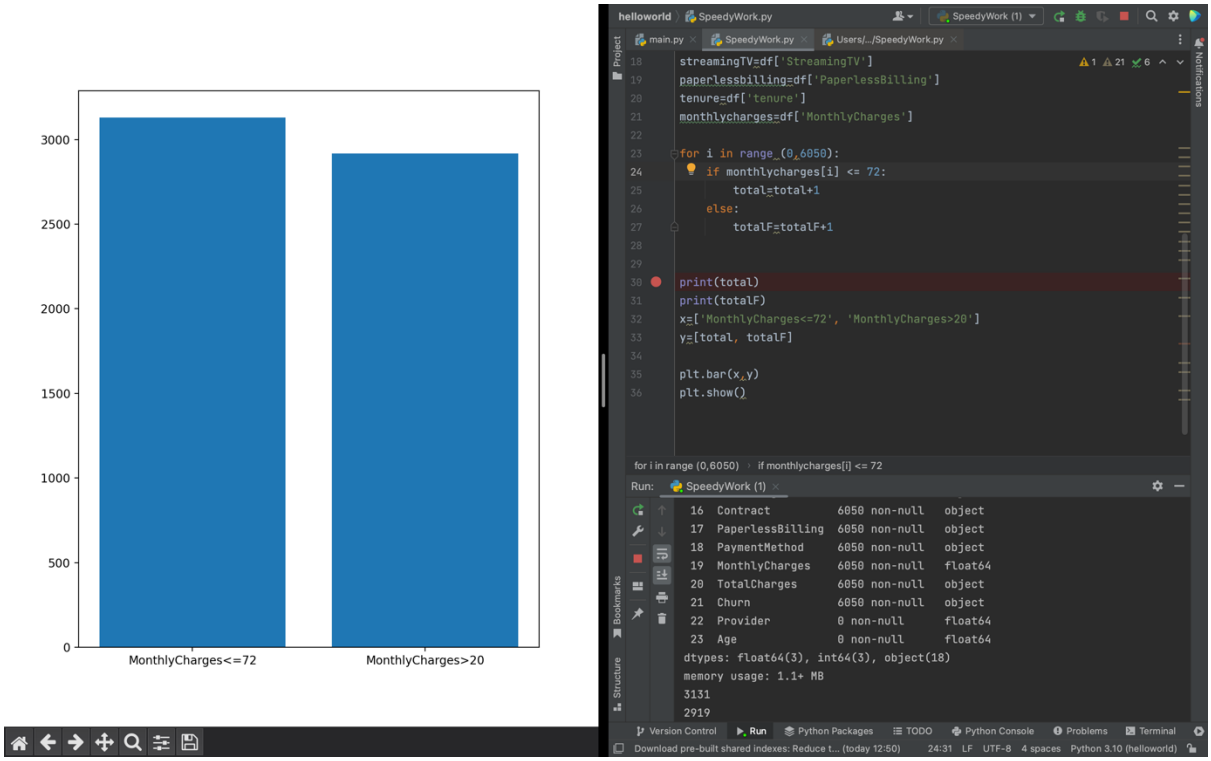
Bar chart comparing the number of people in the data set who had a monthly charge of less than 70 and everyone else in the data.



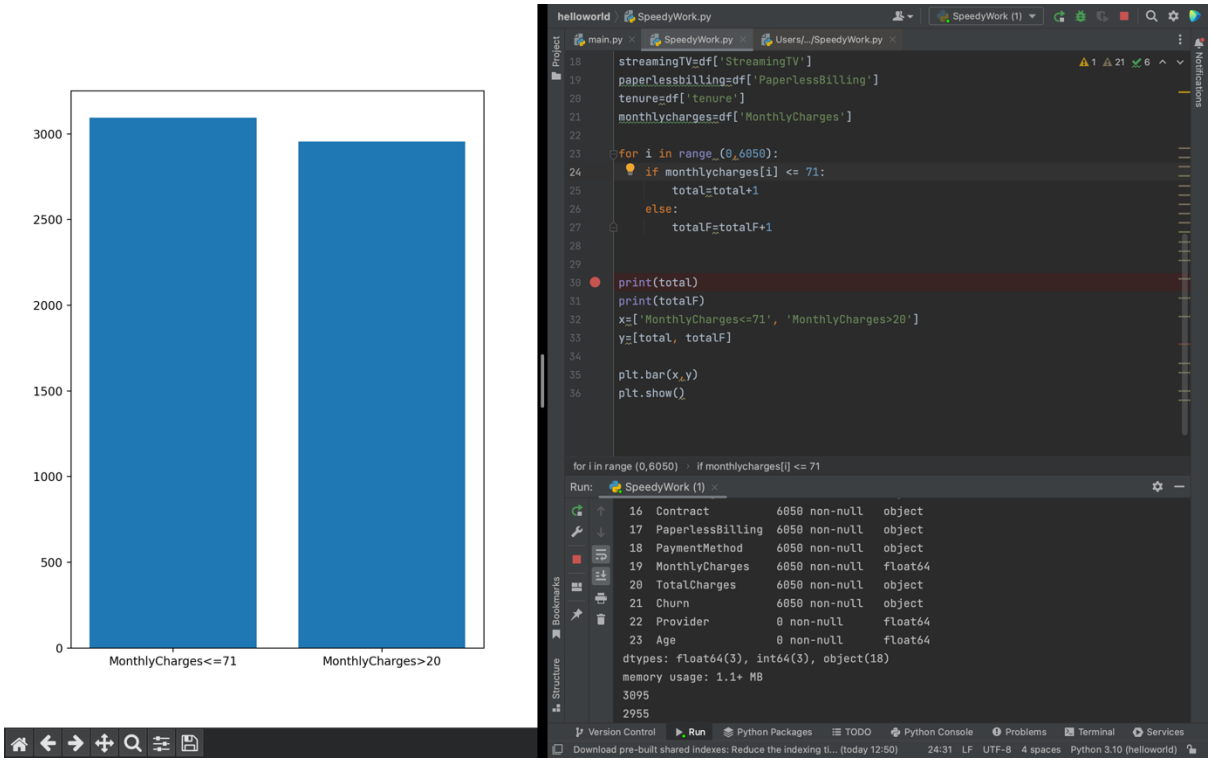
Bar chart comparing the number of people in the data set who had a monthly charge of less than 80 and everyone else in the data.



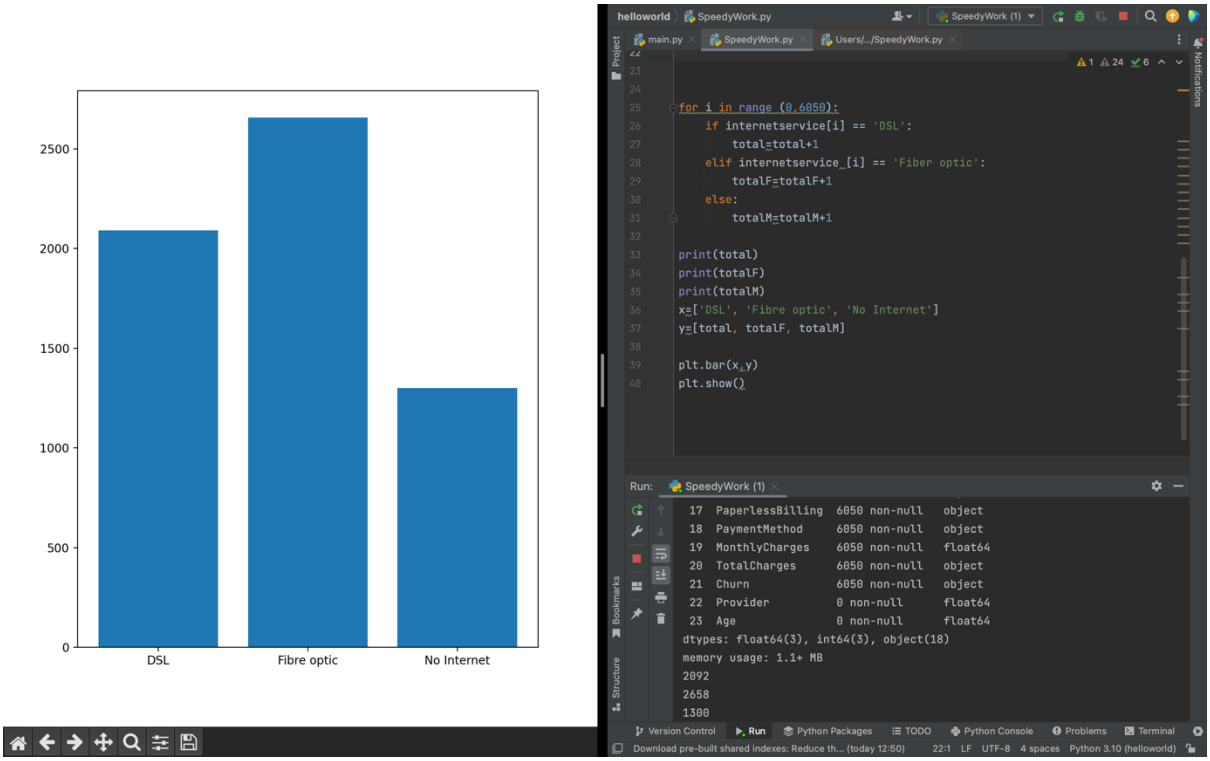
Bar chart comparing the number of people in the data set who had a monthly charge of less than 75 and everyone else in the data.



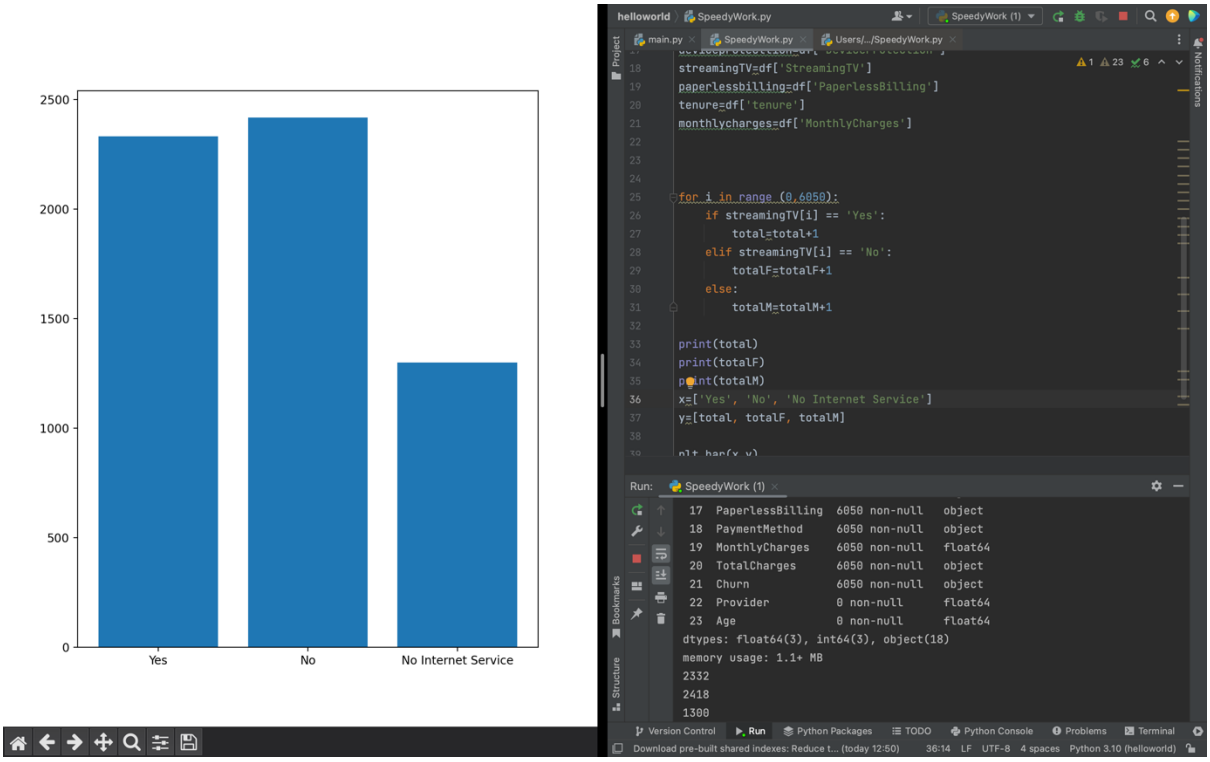
Bar chart comparing the number of people in the data set who had a monthly charge of less than 72 and everyone else in the data.



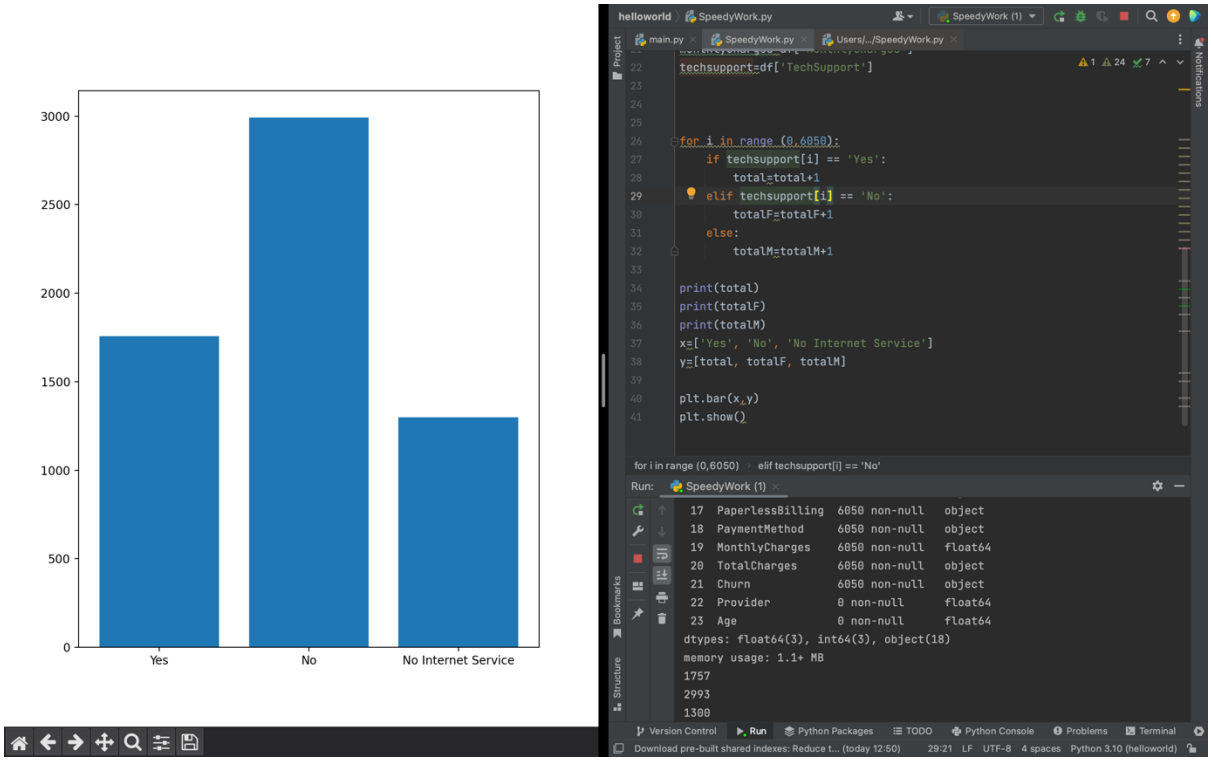
Bar chart comparing the number of people in the data set who had a monthly charge of less than 71 and everyone else in the data. The point at which the majority of people are under is between 70 and 71.



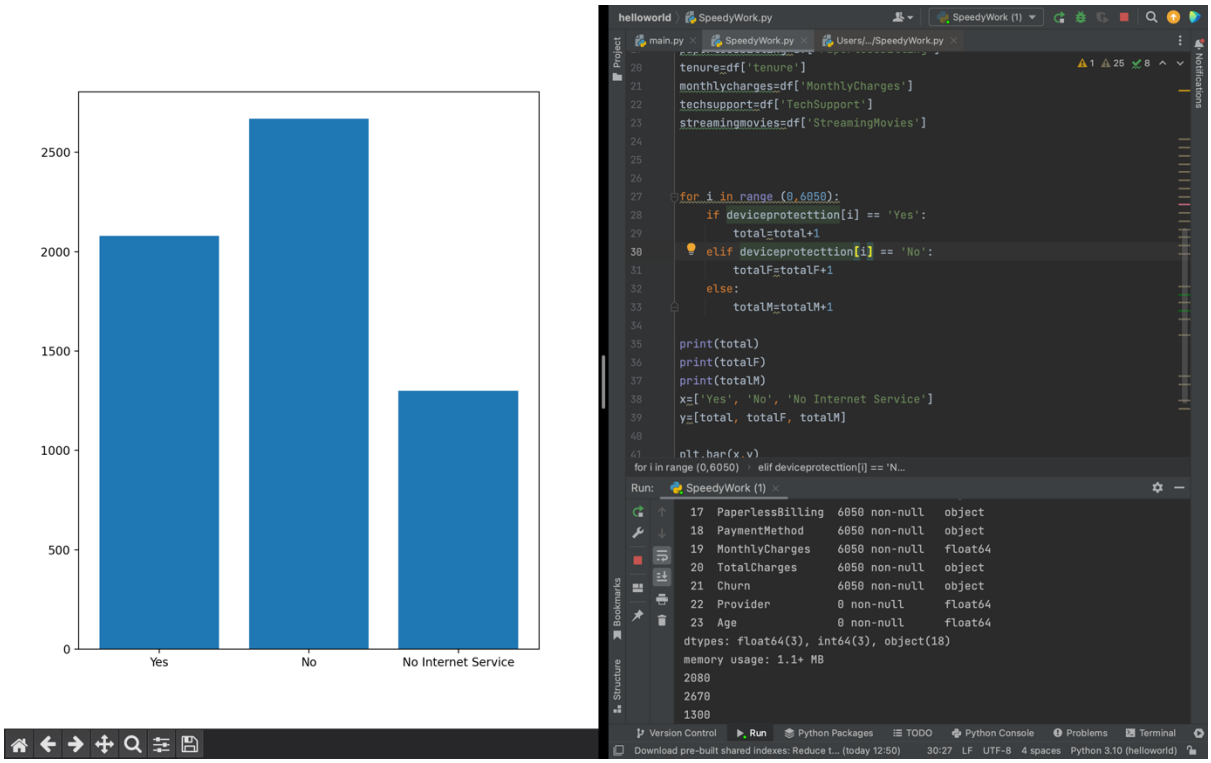
Bar chart displaying the varying responses to if the people in the data had internet service.



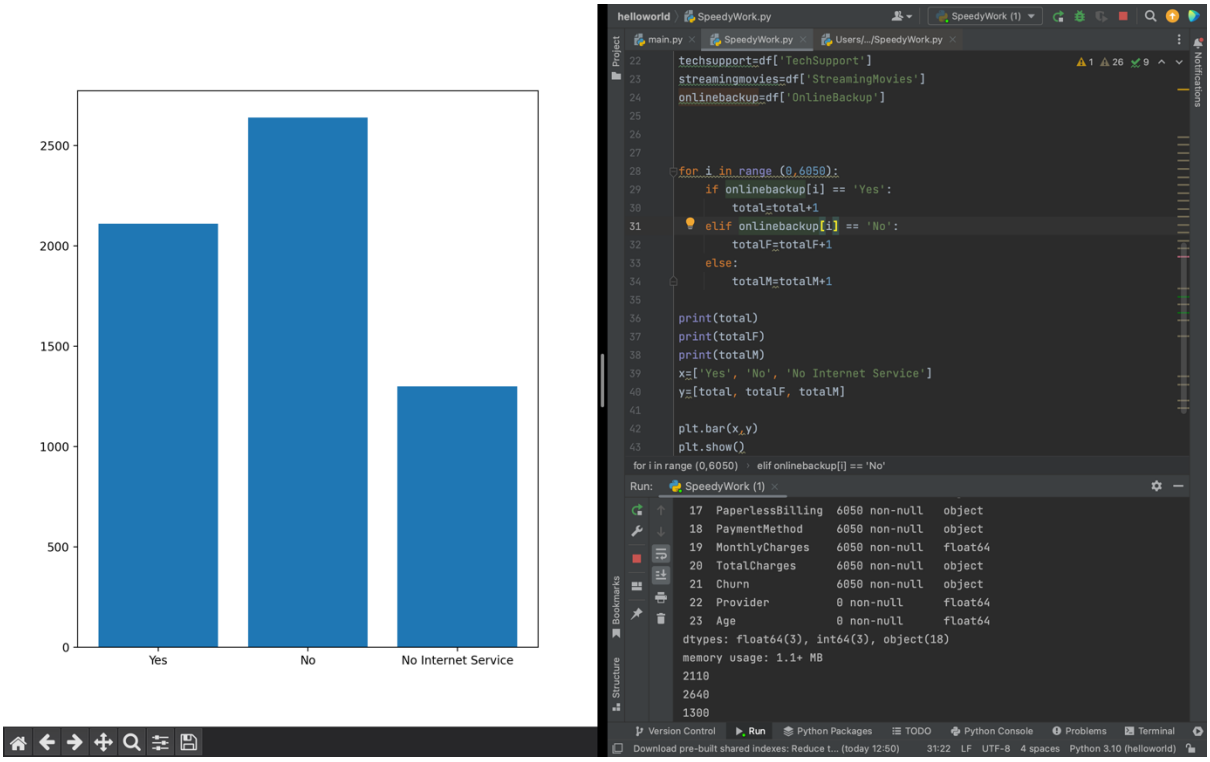
Bar chart displaying the varying responses to if the people in the data had a streaming TV in their homes.



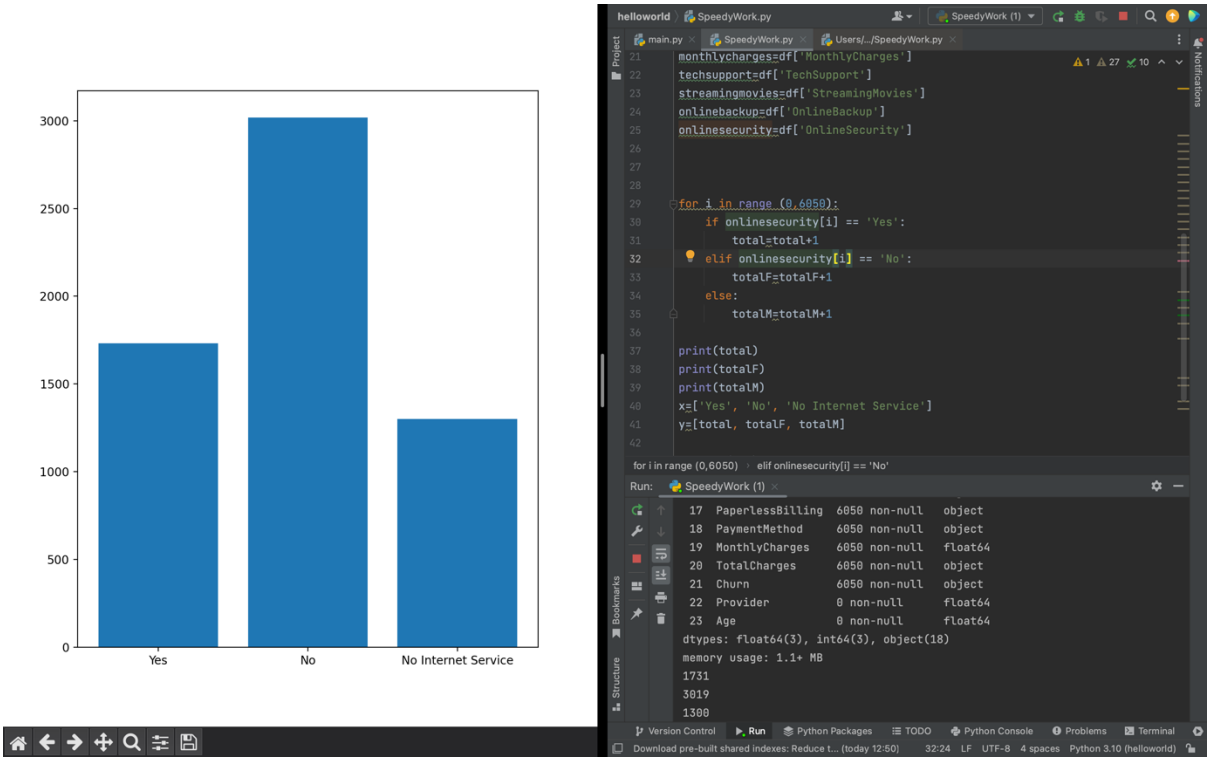
Bar chart displaying the varying responses to if the people in the data had tech support on their devices.



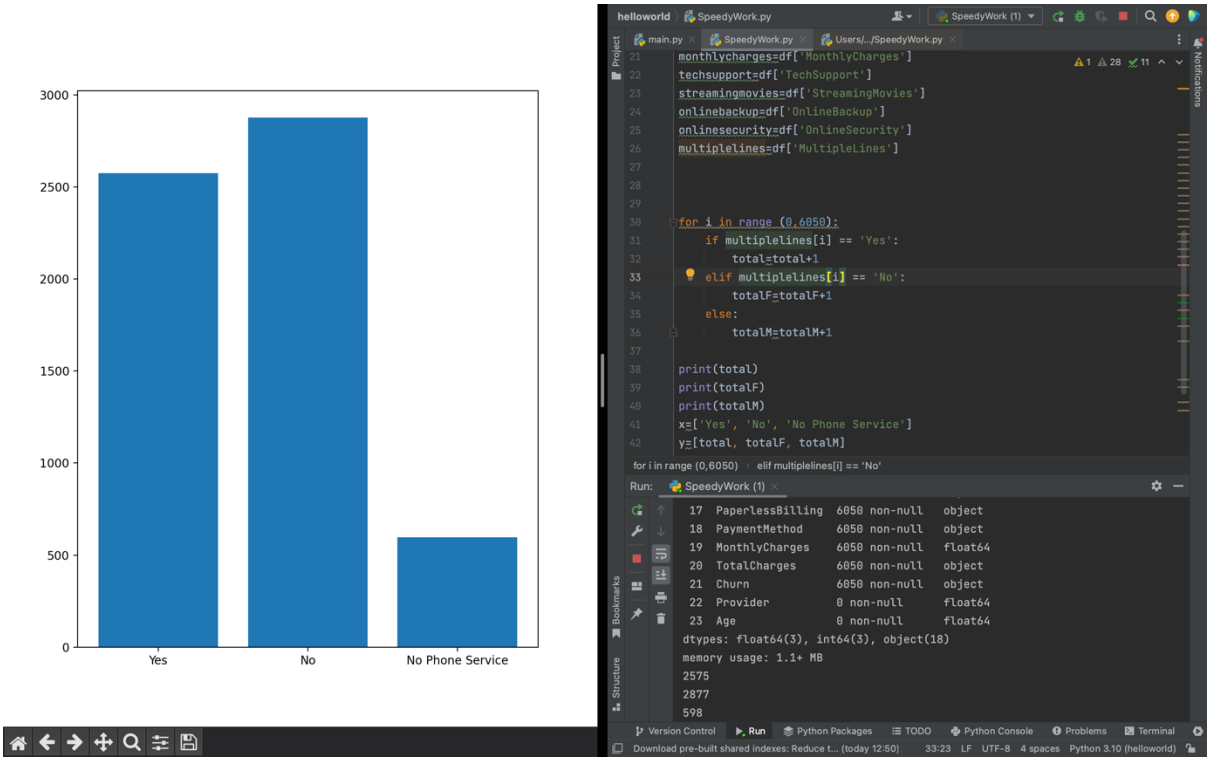
Bar chart displaying the varying responses to if the people in the data had device protection.



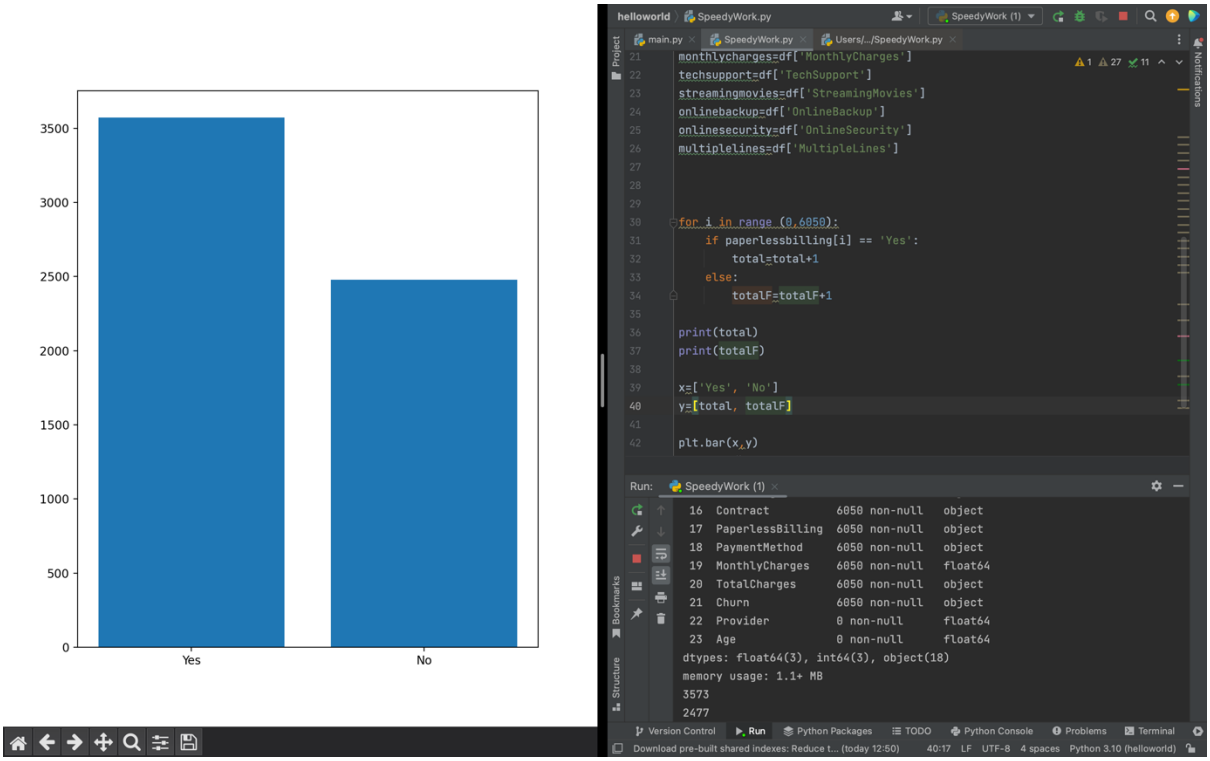
Bar chart displaying the varying responses to if the people in the data had online backup.



Bar chart displaying the varying responses to if the people in the data had online security.



Bar chart displaying the varying responses to if the people in the data had multiple lines.



Bar chart comparing the number of people who have paperless billing.