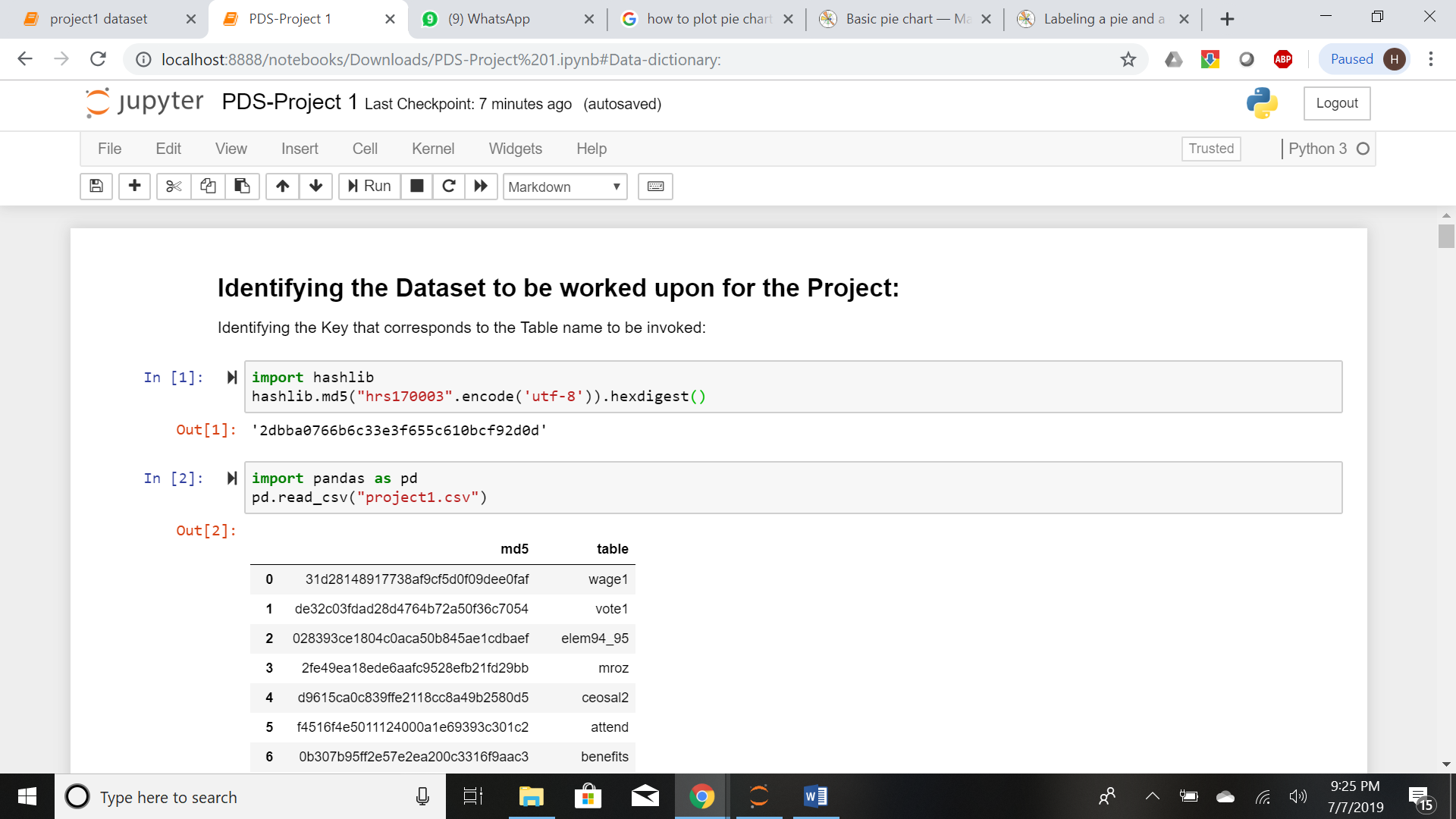
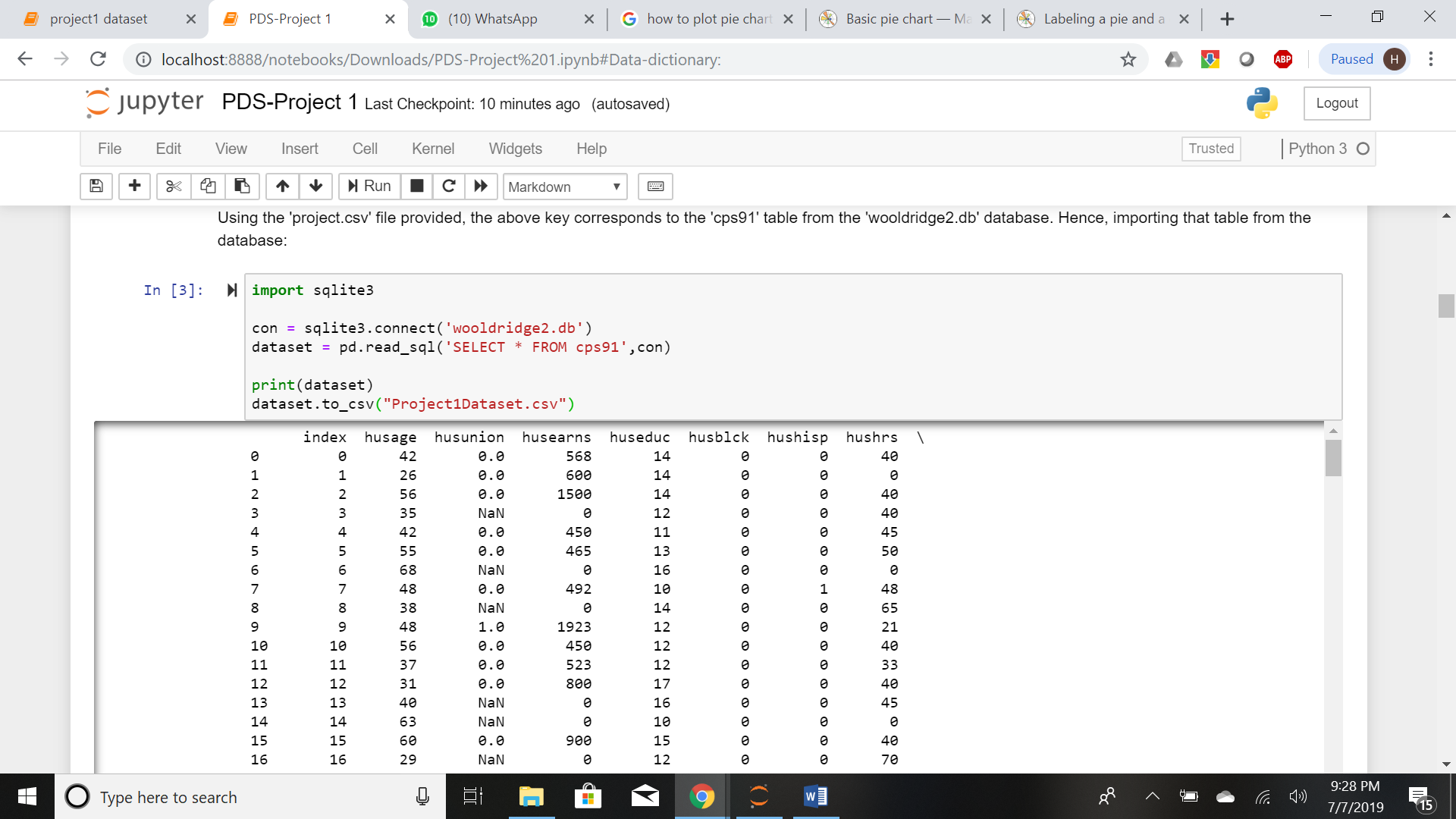
**Project 1 Report**

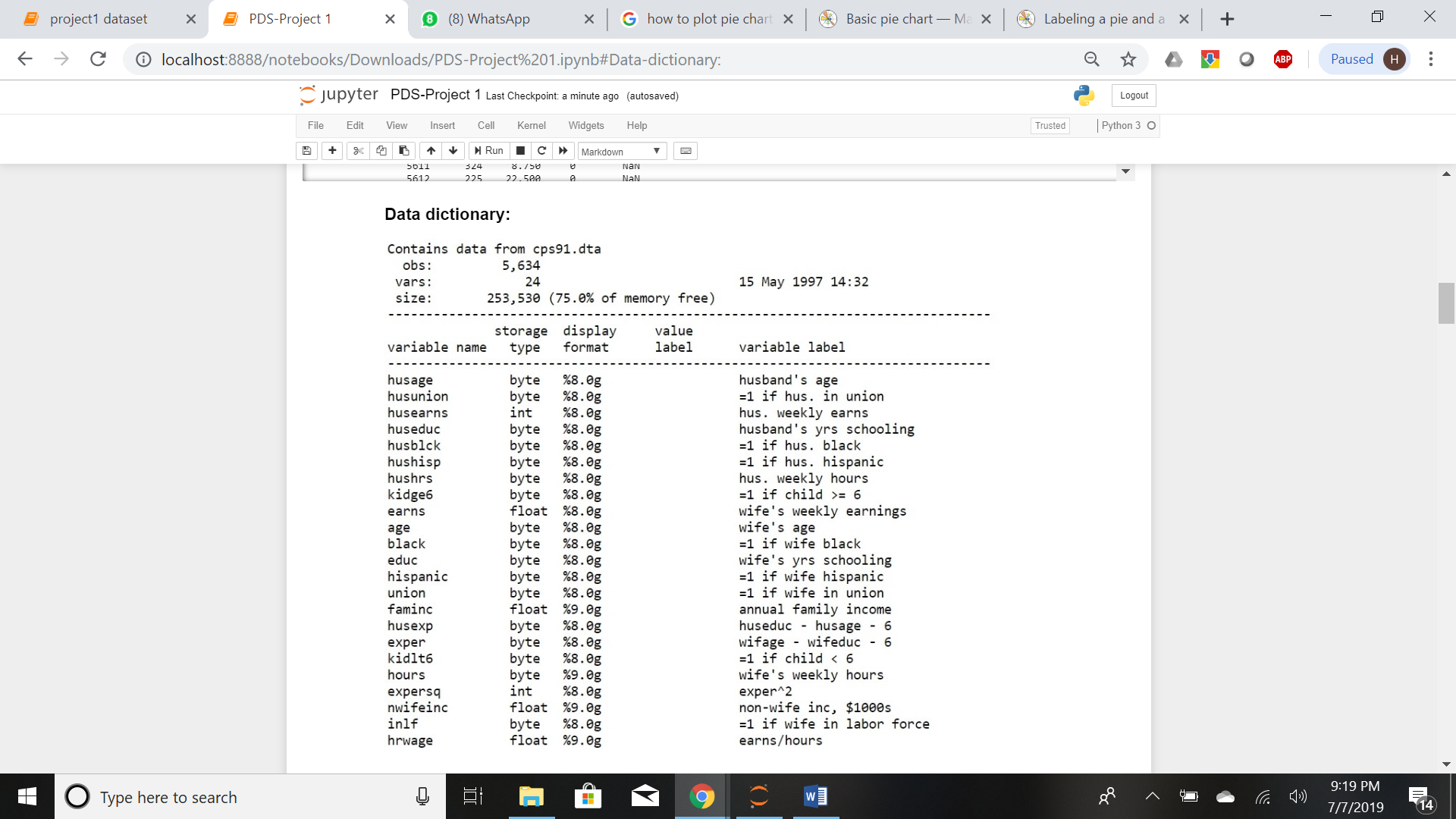
**Step 1: Identifying the Dataset to be worked upon for the Project:**



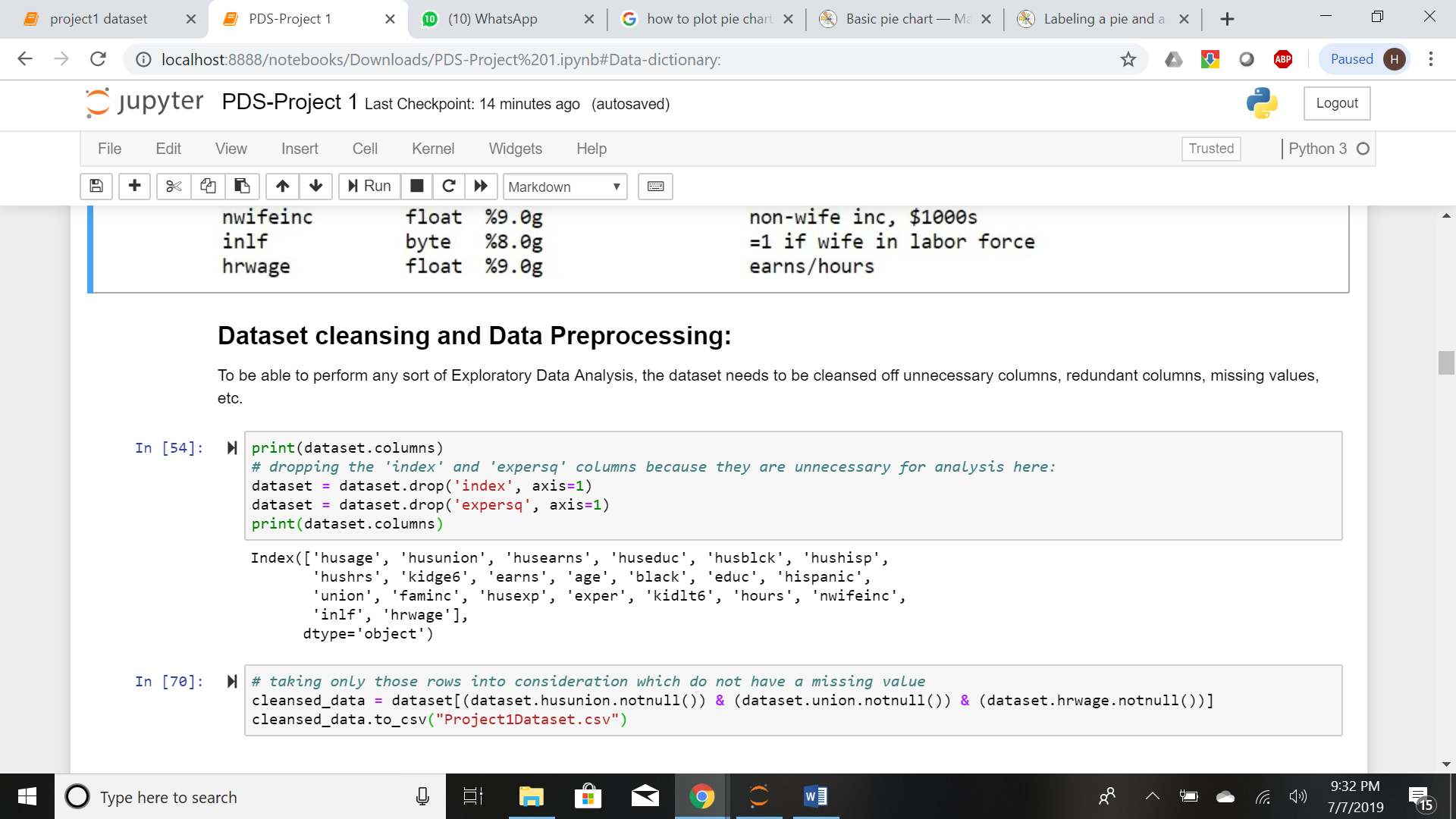
**Step 2: Establishing connection to the Database and accessing the Dataset:**



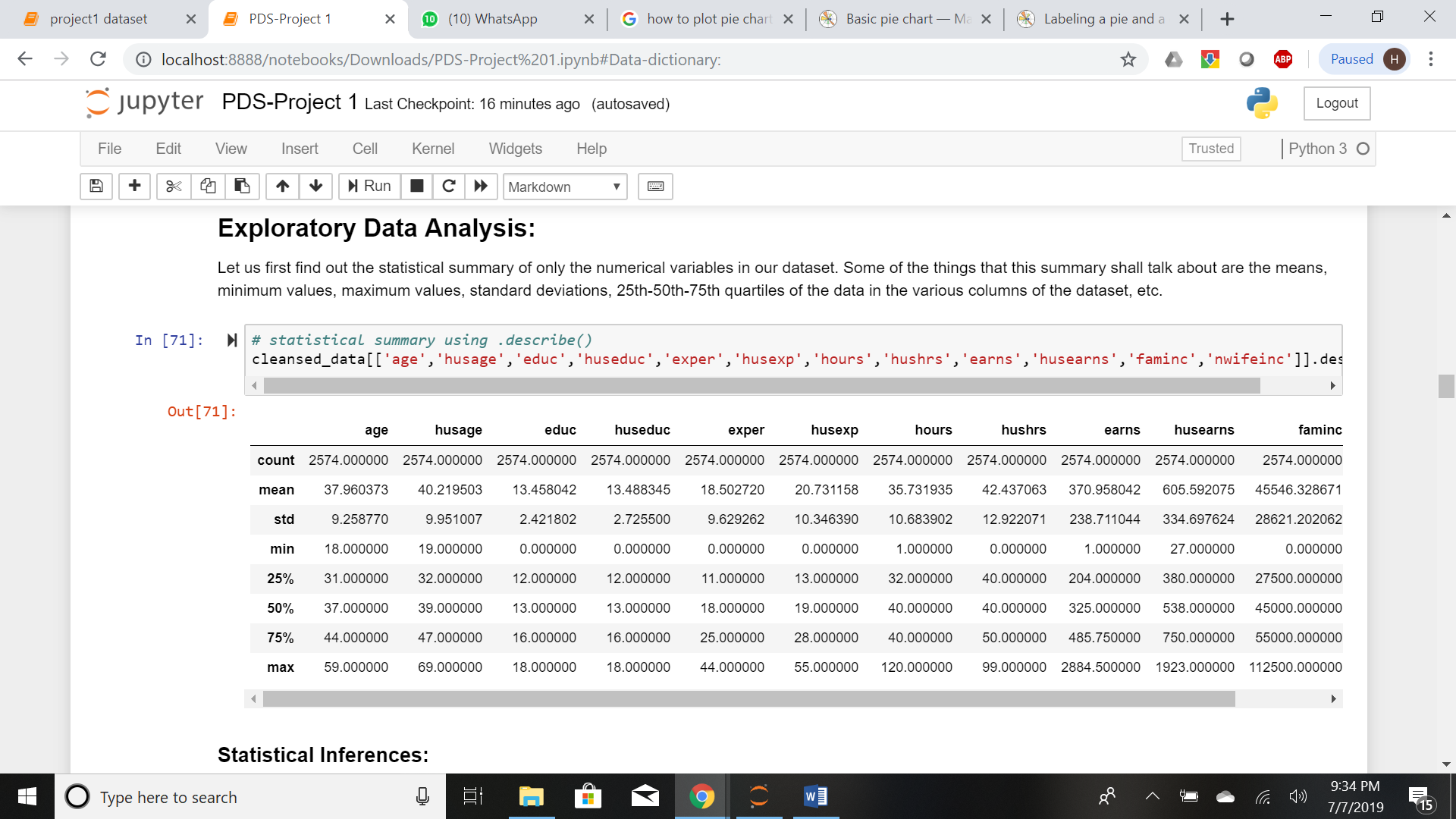
**Step 3: Data Dictionary for the Project:**



**Step 4: Dataset Cleansing and Data Preprocessing:**



**Step 5: Exploratory Data Analysis:**



**Statistical Inferences:**

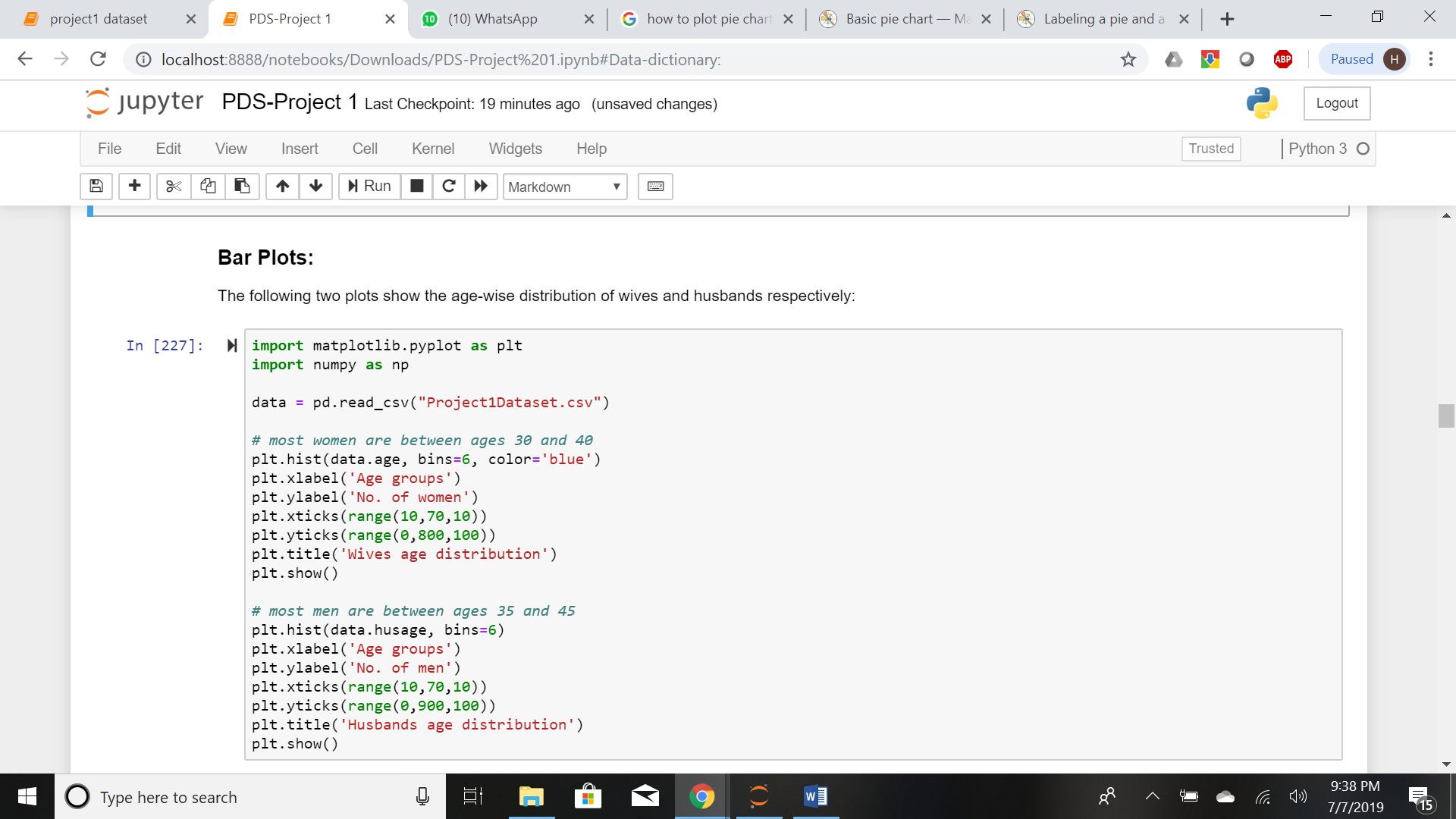
The following inferences are made from the above statistical summary:

1. The mean values, minimum values and maximum values of age of a wife and a husband is not very different.

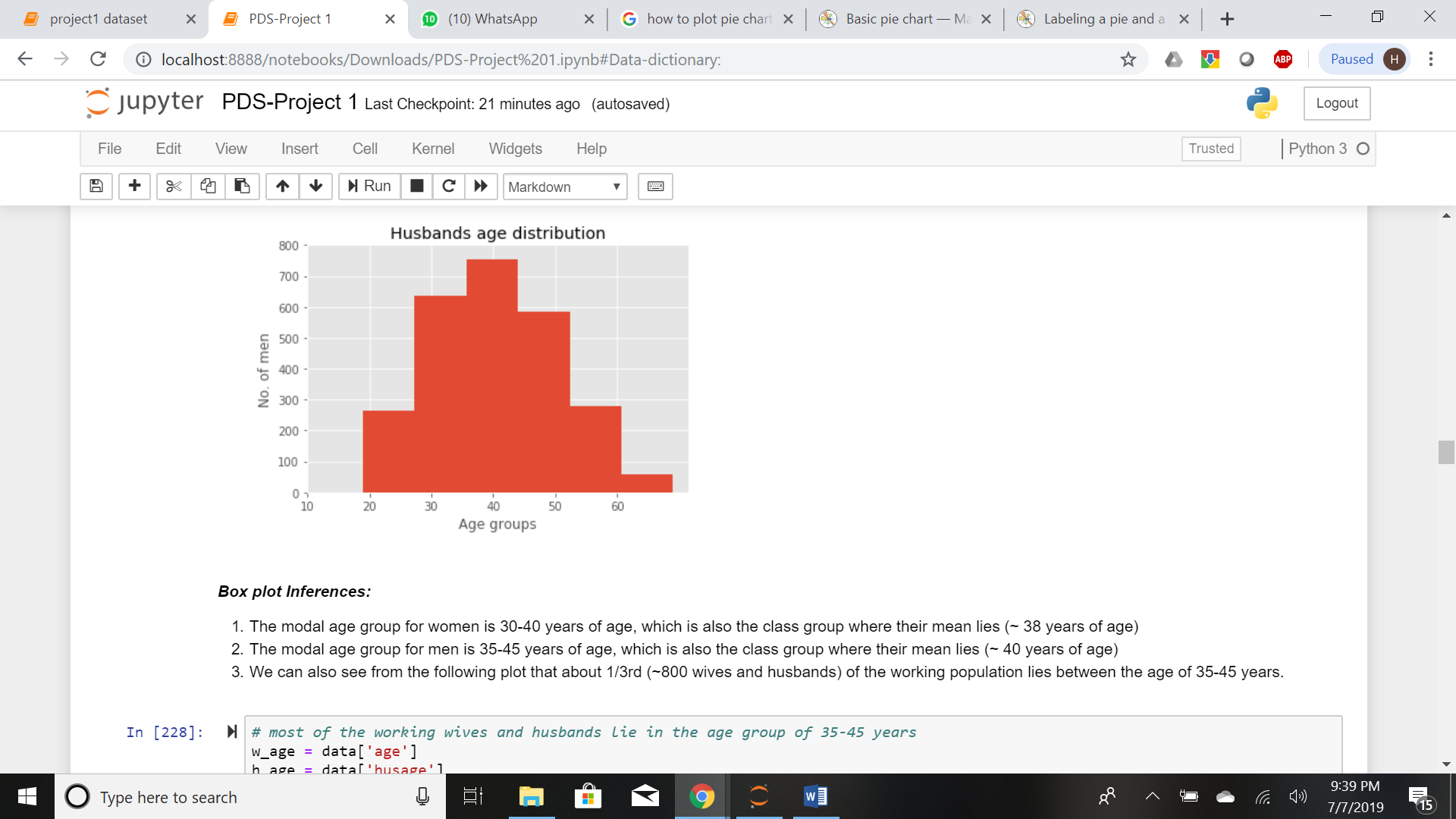
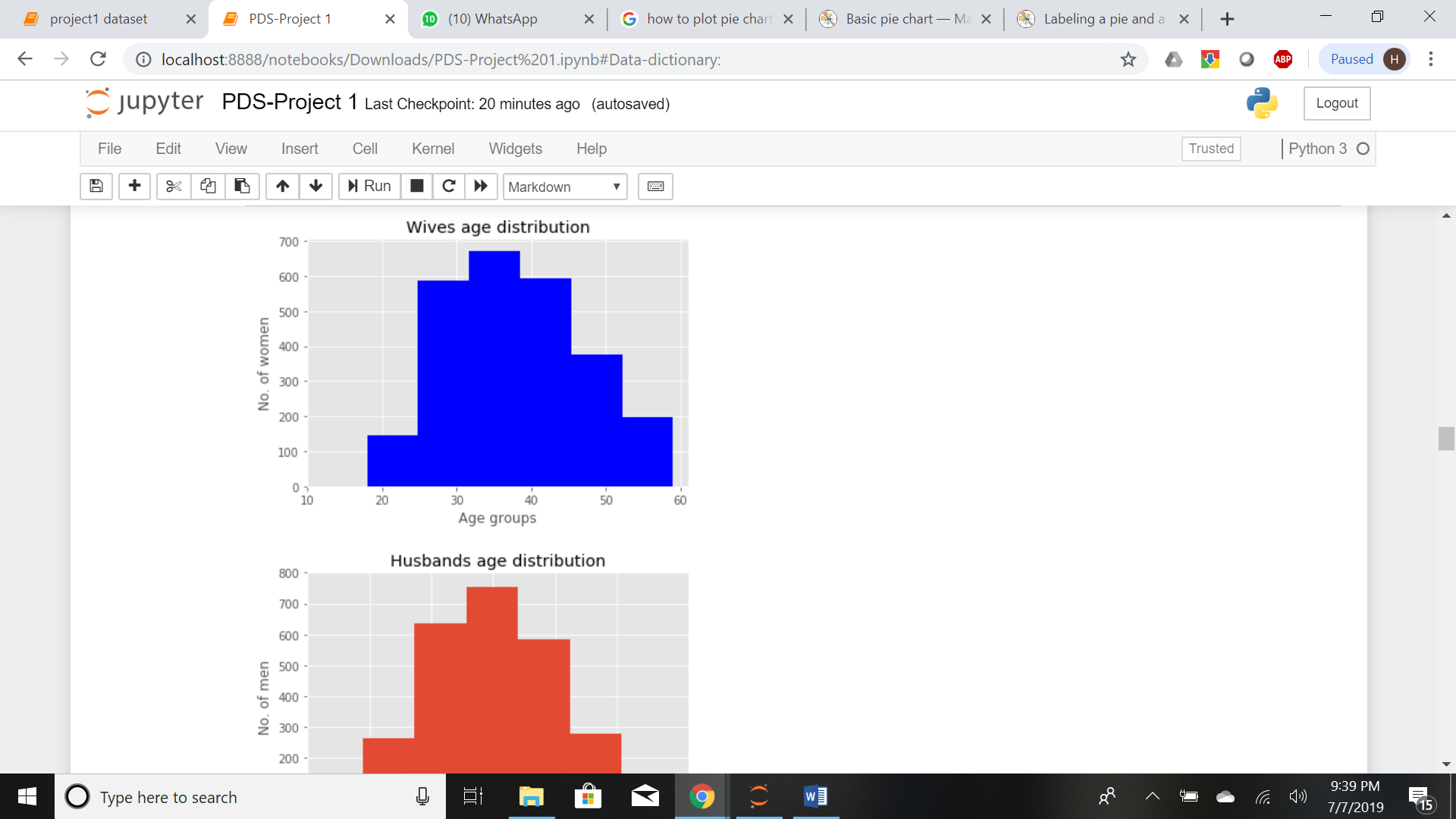
2. Even though the wife and the husband have the same average number of years of educations and almost the same years of experience, the wife earns far less than what the husband earns weekly. This may be because a husband works more hours in a week than a wife does.

3. A non-wife earns far more annually than a wife does.

**Step 6: Bar Plots:**



**Outputs:**

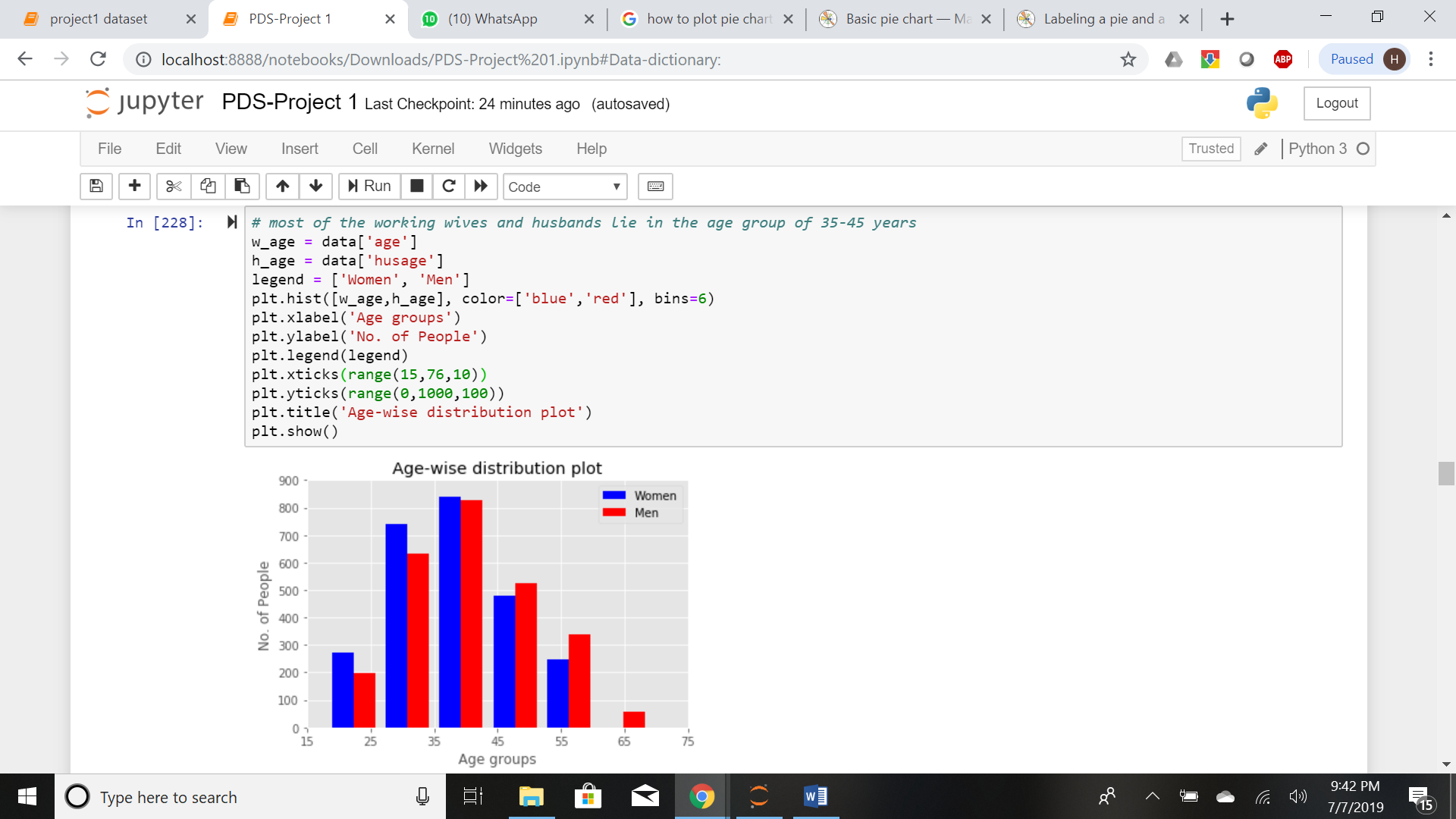


**Box plot Inferences:**

1. The modal age group for women is 30-40 years of age, which is also the class group where their mean lies (~ 38 years of age)

2. The modal age group for men is 35-45 years of age, which is also the class group where their mean lies (~ 40 years of age)

3. We can also see from the following plot that about 1/3rd (~800 wives and husbands) of the working population lies between the age of 35-45 years.



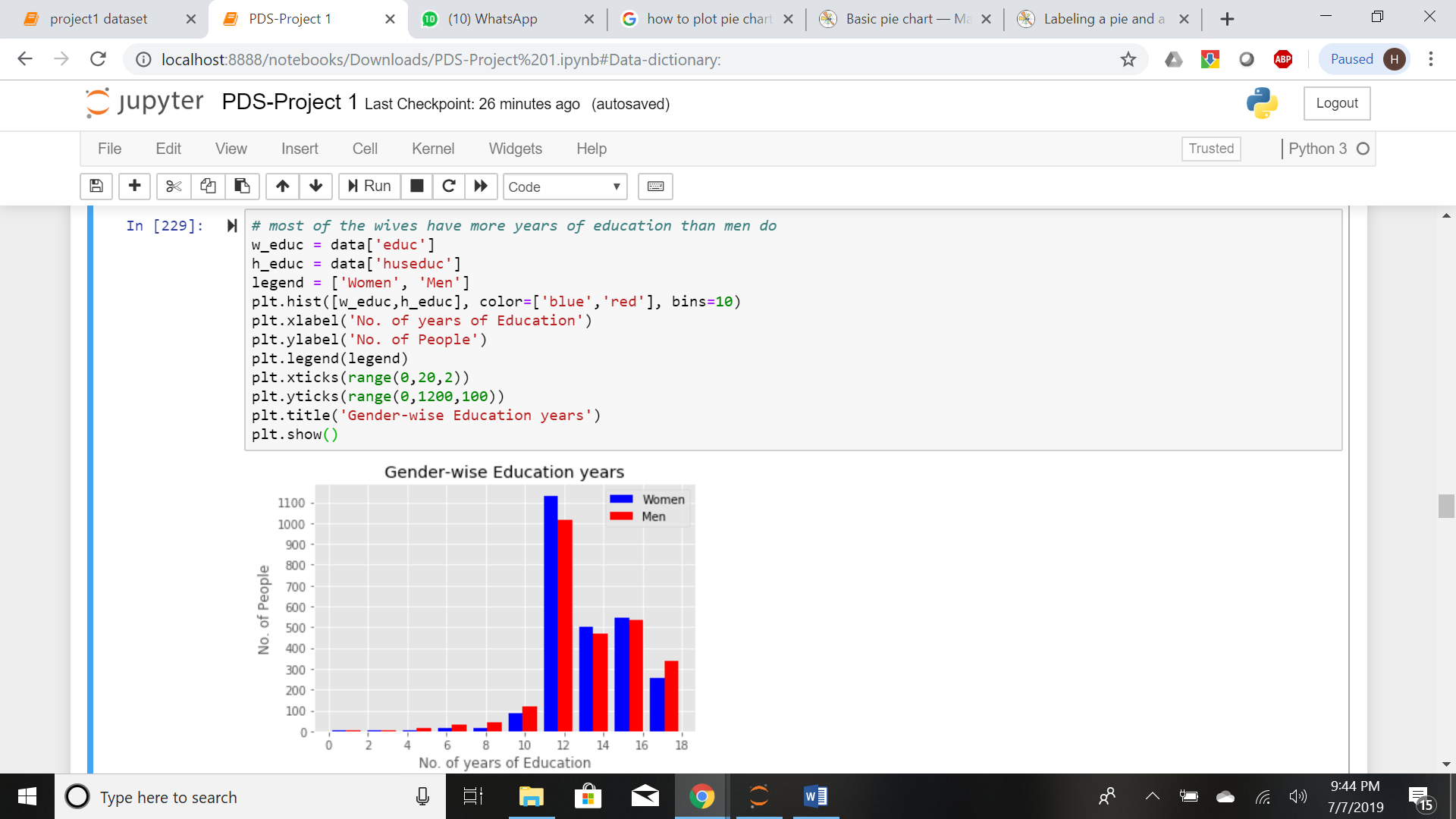
**Step 7: Comparative Analysis and Inference:**

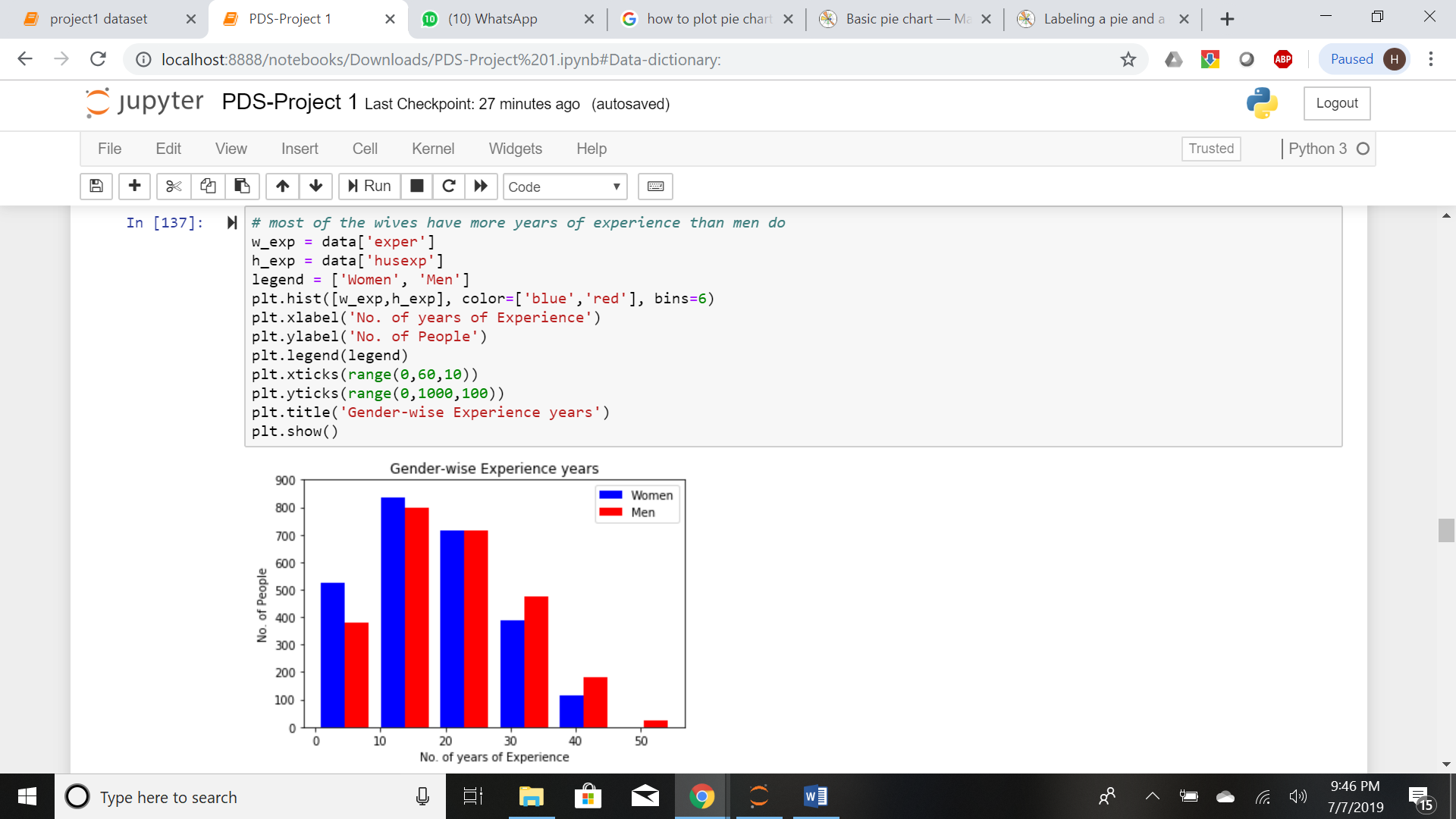
- The following plots of comparison show that most women have obtained more years of education than men.

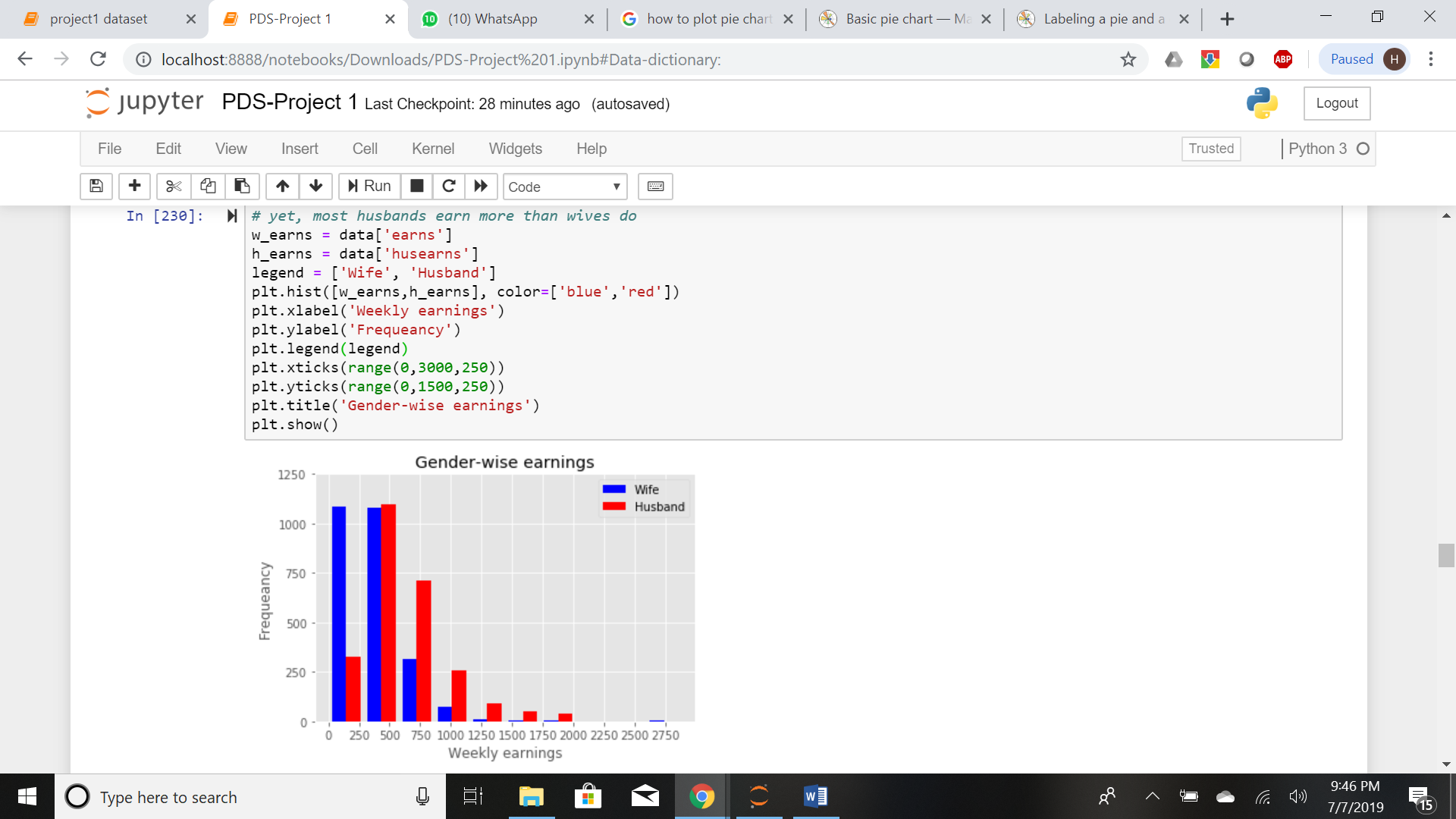
- They also have equal or more experience than men.

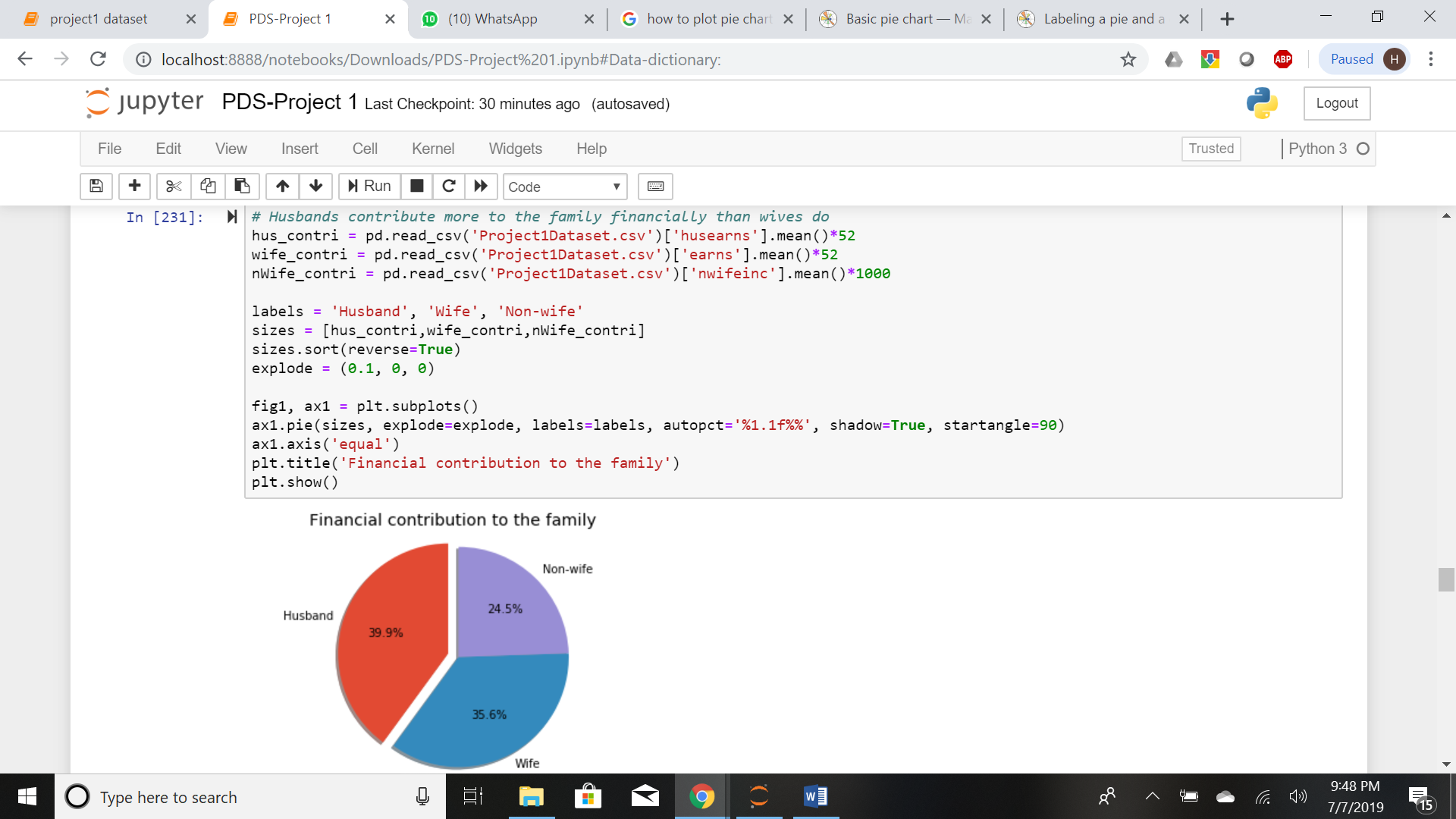
- Yet, most men earn more than women do.

- Also, husbands contribute more to the family financially than wives do.



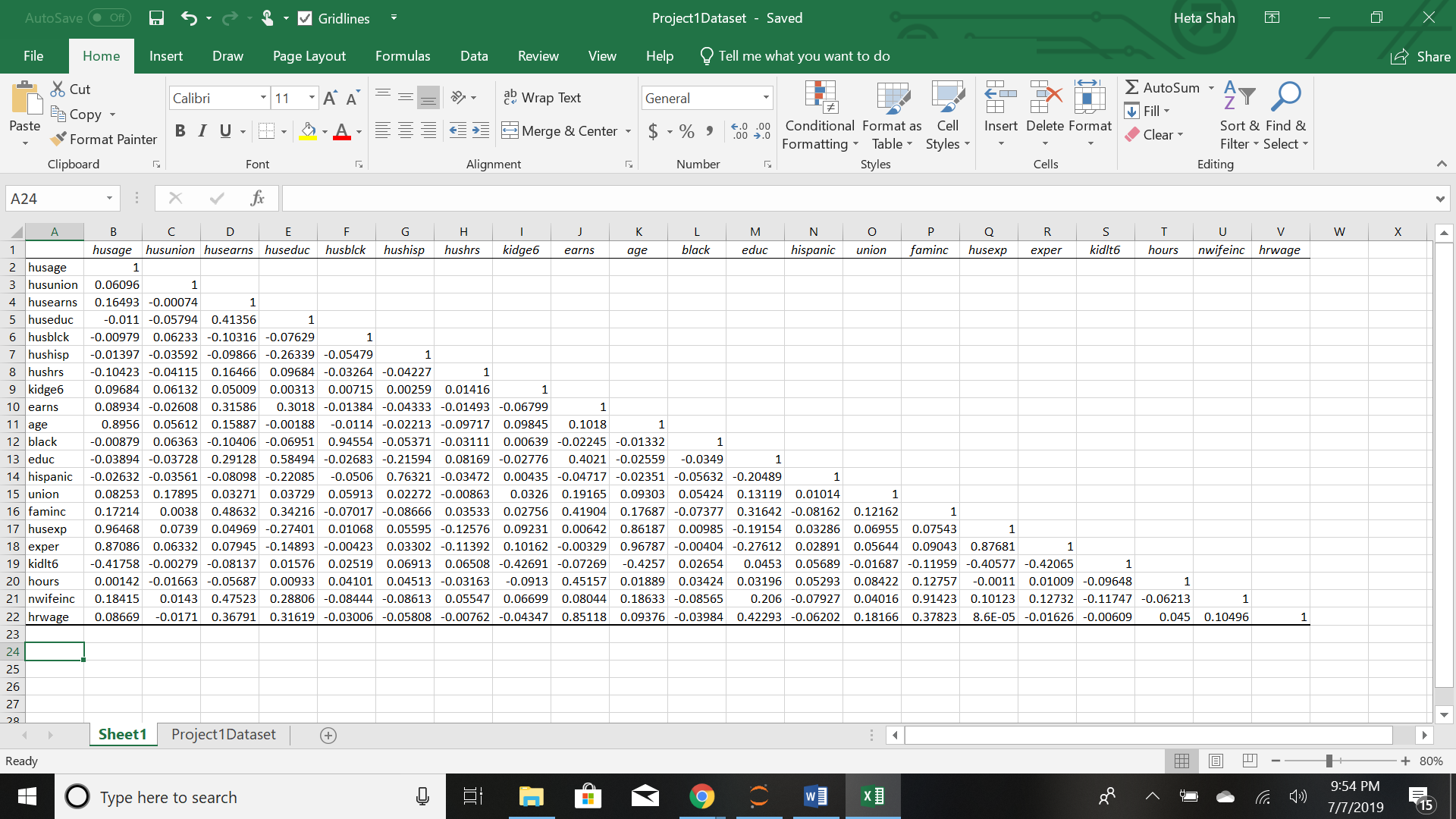






**Step 8: Correlation Matrix:**





**According to the correlation matrix above, the following inferences are made:**

1. There is a strong relation between the wife's and the husband's age (~0.89-0.90). This indicates that women generally get married to men who are close to their age.

2. There is a strong relation between the race of the wives and their husbands. This means that women generally marry men who belong to their race:

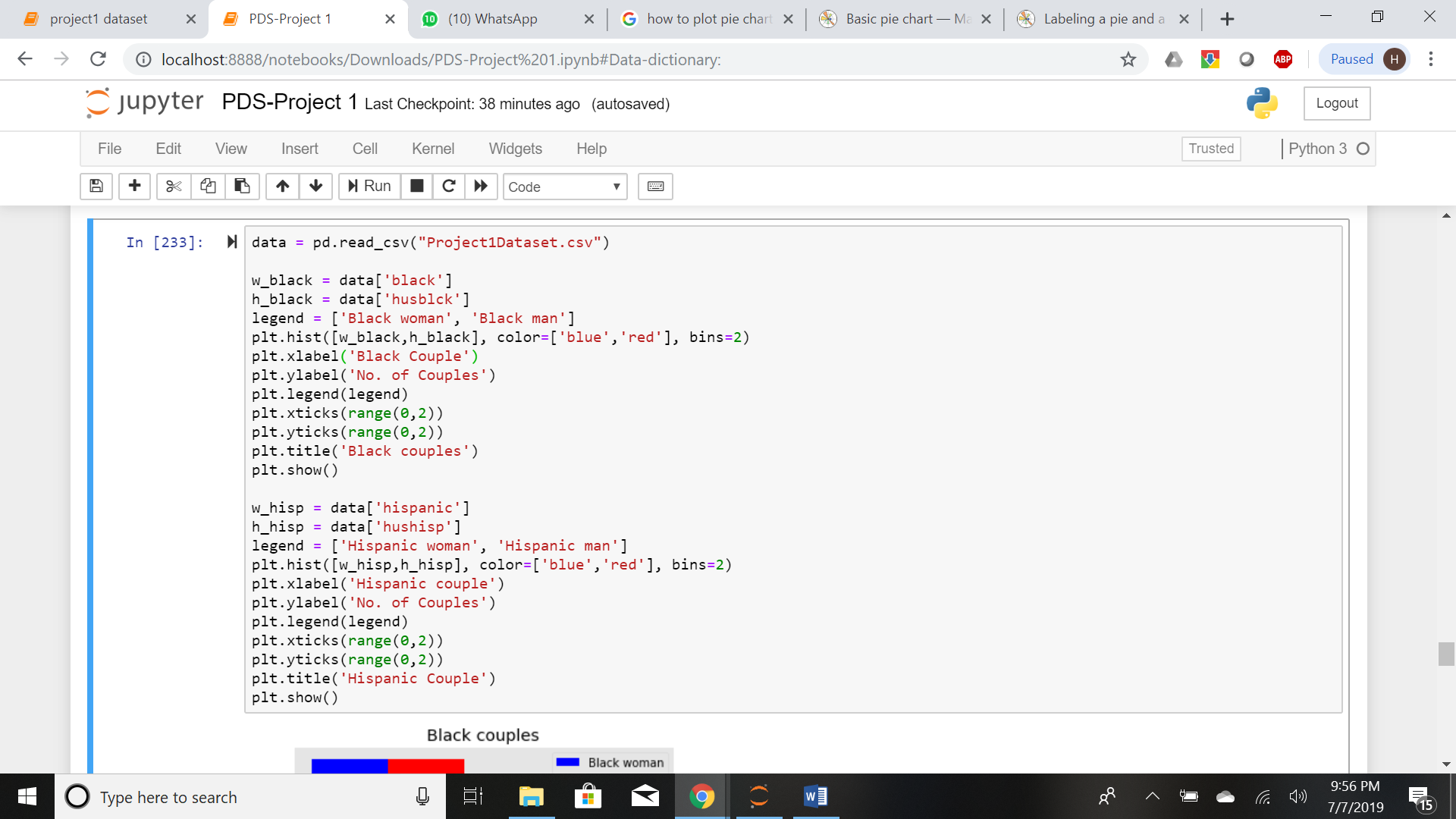
- black couple (~ 0.95)

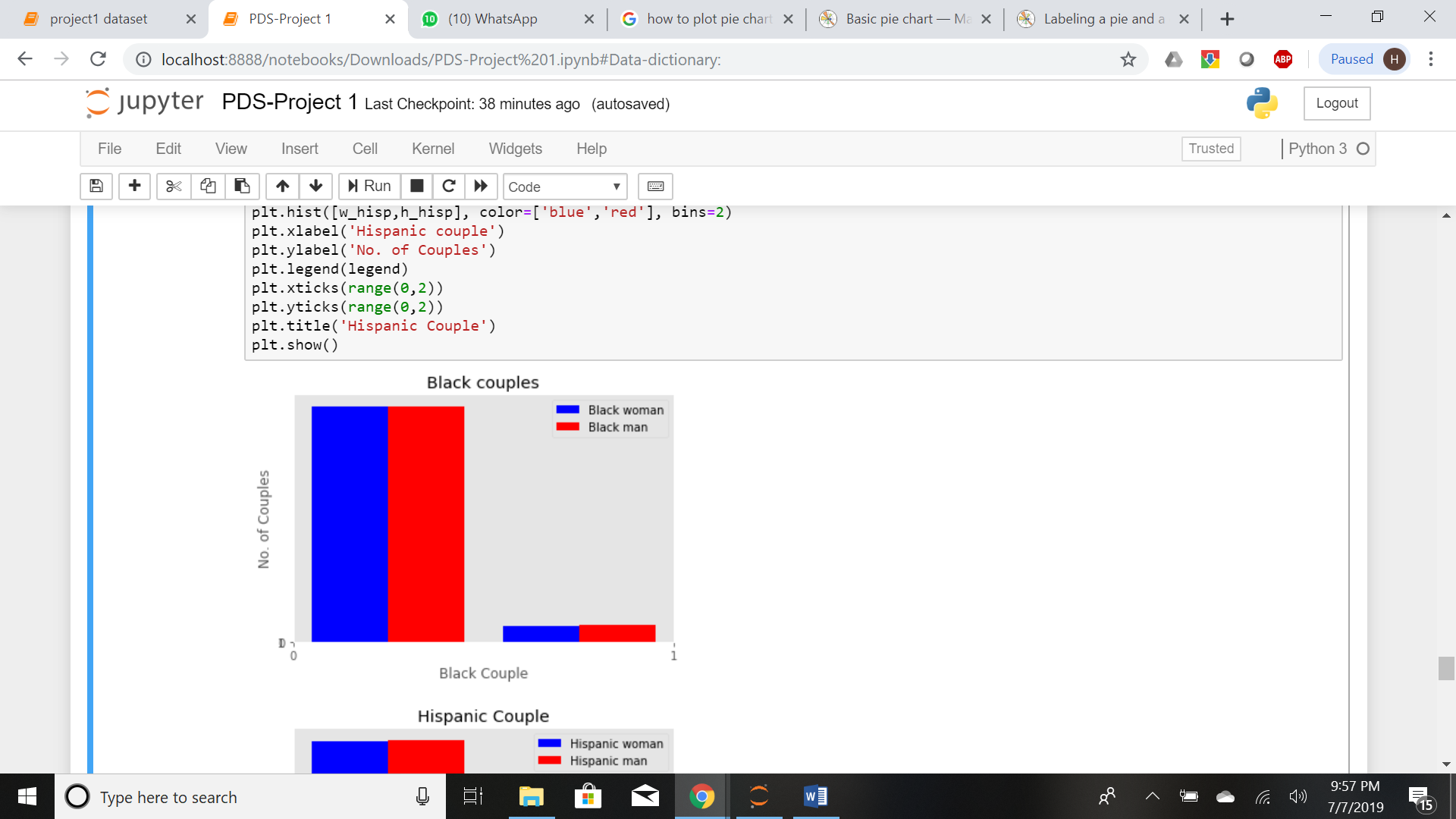
- hispanic couple (~ 0.76)

(refer graph plots given below)

3. The couple's experience and their age show a direct relation (~ 0.97).

4. There is a strong relationship between the family income (wife's income + husband’s income) and the non-wife's income (~ 0.91). This indicates that those families where the couple's income is low/zero, are financially dependent on the non-wife's income.

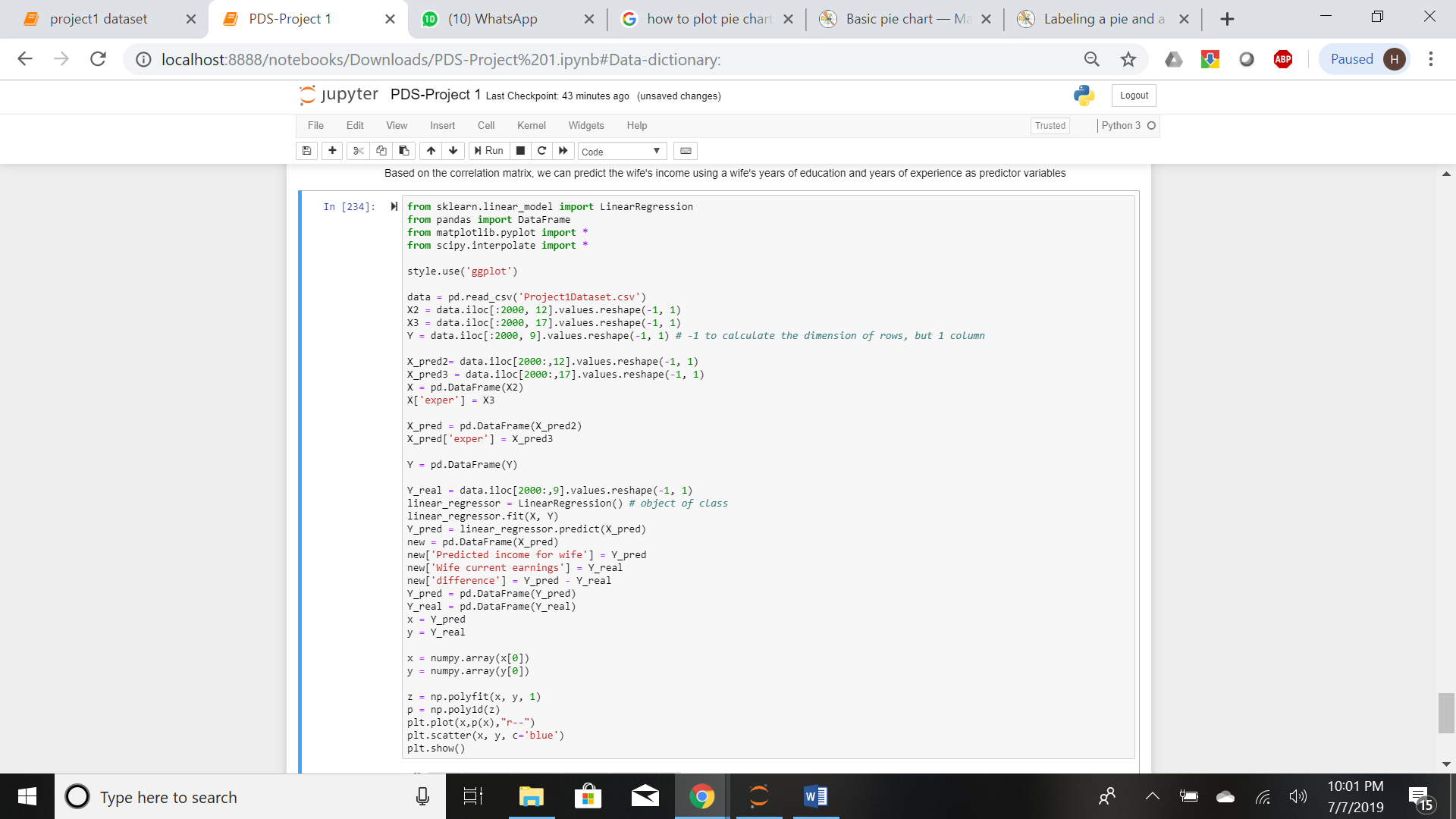


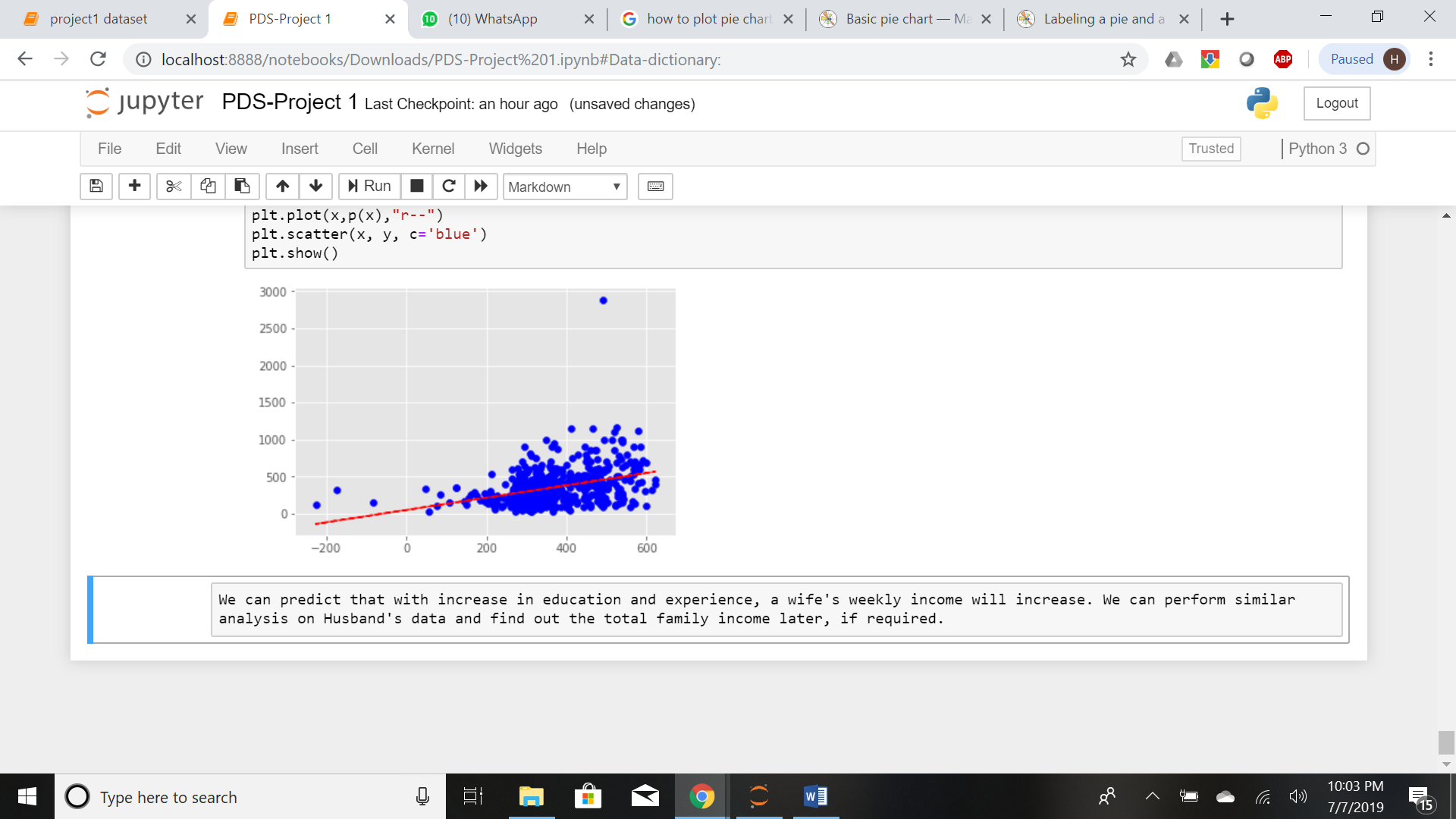


**It can be seen clearly that women usually marry men who belong to their same race.**

**Step 9: Linear Regression to Predict Wife’s Income:**

Based on the correlation matrix, we can predict the wife's income using a wife's years of education and years of experience as predictor variables.





We can predict that with increase in education and experience, a wife's weekly income will increase. We can perform similar analysis on Husband's data and find out the total family income later, if required.

**Step 10: Key Insights:**

1. Even though a wife and a husband have the same average number of years of educations and almost the same average number of years of experience, the wife earns far less than what the husband earns weekly, on average. This may be because a husband works more hours in a week than a wife does.

2. A non-wife earns far more annually than a wife does. This may be because a wife might have to look after the husband and the kid, if any.

3. The modal age group for women is 30-40 years of age, which is also the class group where their mean lies (~ 38 years of age).

4. The modal age group for men is 35-45 years of age, which is also the class group where their mean lies (~ 40 years of age).

5. About 1/3rd (~800 wives and husbands) of the working population lies between the age of 35-45 years.

6. Husbands contribute more to the family financially than wives do.

7. Well-educated women generally get married to well-educated men.

8. Women generally marry men who belong to their same race:

9. There is a strong relationship between the family income (wife's income + husband's income) and the non-wife's income (~ 0.91). Those families where the couple's income is low/zero, are financially dependent on the non-wife's income.

10. Using linear regression, we have predicted that with increase in education and experience, a wife's weekly income will increase. We can perform similar analysis on Husband's data and find out the total family income later, if required.