**Internship Report**

(Project Work)

**On**

**INCOME PREDICTION OF A USA DATABASE USING MACHINE LEARNING**

*Submitted to*

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR, ANANTHAPURAMU**

*In Partial Fulfillment of the Requirements for the Award of the Degree of*

**BACHELOR OF TECHNOLOGY**

**In**

**COMPUTER SCIENCE & TECHNOLOGY**

**Submitted By**

## R. ANURAAG - 20691A2805

**Under the Guidance of**

**Mr. Y. RAVI RAJU**

**ASST PROFESSOR**

**Department of Computer Science & Technology**



**MADANAPALLE INSTITUTE OF TECHNOLGY & SCIENCE**

**(UGC – AUTONOMOUS)**

**(Affiliated to JNTUA, Ananthapuramu)**

**Accredited by NBA, Approved by AICTE, New Delhi)**

**AN ISO 9001:2008 Certified Institution**

**P. B. No: 14, Angallu, Madanapalle, Annamayya – 517325**

## DEPARTMENT OF COMPUTER SCIENCE & TECHNOLOGY

**BONAFIDE CERTIFICATE**

This is to certify that the **SUMMER INTERNSHIP-1 (20CST701)** entitled **“**INCOME PREDICTION OF A USA DATABASE USING MACHINE LEARNING**”** is bonafide work carried out by

### R. ANURRAG - 20691A2805

Submitted in partial fulfillment of the requirements for the award of degree **Bachelor of Technology** in the stream of **Computer Science & Technology** in **Madanapalle Institute of**

**Technology & Science, Madanapalle,** affiliated to **Jawaharlal Nehru Technological University Anantapur, Ananthapuramu** during the academic year 2022-2023

Guide Head of the Department

**Mr. Y. RAVI RAJU Dr. M. Sreedevi**

**Assistant Professor, Professor and Head**

**Department of CST Department of CST**

**INTERNSHIP CERTIFICATE:**



### DECLARATION

We hereby declare that the results embodied in this **SUMMER INTERNSHIP-1**

**(20CST701) “** INCOME PREDICTION OF A USA DATABASE USING MACHINE LEARNING**”** by us under the guidance of **Mr. Y. RAVI RAJU, Assistant Professor, Dept. of CST** in partial fulfillment of the award of **Bachelor of Technology** in **Computer Science & Technology** from **Jawaharlal Nehru Technological University Anantapur, Ananthapuramu** and we have not submitted the same to any other University/institute for award of any other degree.

**Date:**

**Place:**

**PROJECT MEMBER**

### 

**R. ANURAAG - 20691A2805**

I certify that above statement made by the students is correct to the best of my knowledge.

**Date:**

**Guide:**

## TABLE OF CONTENT

**S.NO TOPIC PAGE NO.**

### 1. INTRODUCTION 8

1.1 About Industry or Organization Details 8

1.2 My Personal Benefits 8

1.3 Objective of the Project 8

1.4 Limitations of Project 9

### 2. SYSTEM ANALYSIS 10

2.1 Introduction 10

2.2Proposed System 11

2.3 Advantages over Existing System 12

### 3. SYSTEM SPECIFICATION 13

3.1 Hardware Requirement Specification 13

3.2 Software Requirement Specification 13

### 4. SYSTEM DESIGN 14

4.1 System Architecture 15

4.2 Modules Flow diagrams 16

### 5. IMPLEMENTATION AND RESULTS 17

5.1 Implementation 17-27

1. **CONCLUSION 28**

7.1 Conclusion 28

**REFERENCES 28**

**ABSTRACT**

Machine Learning is one of the emerging fields of Artificial Intelligence and it has many applications. It is a tool that uses data and Artificial Intelligence in its areas of application. The main idea behind the development of machine learning algorithms is to create a model that understands and analyzes the given data and helps in prediction. Machine learning methods can be applied to various domains.

Principal component analysis and support vector machine methods are employed to generate and evaluate income prediction data based on the Current Population Survey provided by the U.S. Census Bureau. A detailed statistical study targeted for relevant feature selection is found to increase efficiency and even improve classification accuracy. A systematic study is performed on the influence of this statistical narrowing on the grid parameter search, training time, accuracy, and number of support vectors. Accuracy values as high as 84%, when compared against a test population, are obtained with a reduced set of parameters while the computational time is reduced by 60%. Tailoring computational methods around specific real data sets is critical in designing powerful algorithms.

**Keywords:** Support vector machine, principal component analysis, classification, accuracy.

**CHAPTER 1** [](https://www.smartknower.com/)

**INTRODUCTION**

* 1. **About Industry or Organization Details:**

* Smart Knower is a edu-tech learning space with AI base software at its core   
  which is located in –Bengaluru, INDIA.
* Smart knower offers a connected ecosystem accessible from anywhere and by anyone.
* It provides Internships to enhance the knowledge in IT industry through Entrepreneurial domains such as Machine learning, Full stack web development, etc.
* It has well equipped resources for internships such as Mentors, LMS portal with contains recorded sessions.
  1. **My Personal Benefits:**
* I have gained some knowledge in Machine learning.
* I have learned the importance of time management.
* I have learned the doing work in the given time.
* I have learned how to work in a group project.
* I have learned how to interact with people in a project group.
  1. **Objective of the Project:**
* To get trained in python from scratch as well as Machine Learning algorithms and to develop basic projects of it.

* 1. **Limitations of Project:**
* Stretched resources.
* Low of clarity.
* Limited time.
* Given datasets.

**CHAPTER 2**

**SYSTEM ANALYSIS**

**2.1 Introduction:**

**History of Machine Learning Machine:**

Learning was first conceived from the mathematical modelling of neural networks. A paper by logician Walter Pitts and neuroscientist Warren McCulloch, published in 1943, attempted to mathematically map out thought processes and decision making in human cognition. In 1950, Alan Turning proposed the Turing Test, which became the litmus test for which machines were deemed “intelligent” or “unintelligent.” The criteria for a machine to receive status as an “intelligent” machine, was for it to have the ability to convince a human being that it, the machine, was also a human being. Soon after, a summer research program at Dartmouth College became the official birthplace of AI. Intelligent machines went on to do everything from using speech recognition to learning to pronounce words the way a baby would learn to defeating a world chess champion at his own game. The infographic below shows the history of machine learning and how it grew from mathematical models to sophisticated technology.

**Machine learning:**

Machine learning (ML) is a field of inquiry devoted to understanding and building methods that ‘learn’, that is, methods that leverage data to improve performance on some set of tasks. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a 10 wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms.

**2.2 PROPOSED SYSTEM:**

**Supervised Learning:**

Supervised learning is analogous to training a child to walk. You will hold the child’s hand, show him how to take his foot forward, walk yourself for a demonstration and so on, until the child learns to walk on his own.

**Regression:**

Similarly, in the case of supervised learning, you give concrete known examples to the computer. You say that for given feature value x1 the output is y1, for x2 it is y2, for x3 it is y3, and so on.Based on this data, you let the computer figure out an empirical relationship between x and y. Once the machine is trained in this way with a sufficient number of data points, now you would ask the machine to predict Y for a given X. Assuming that you know the real value of Y for this given X, you will be able to deduce whether the machine’s prediction is correct. Thus, you will test whether the machine has learned by using the known test data. Once you are satisfied that the machine is able to do the predictions with a desired level of accuracy (say 80 to 90%) you can stop further training the machine. Now, you can safely use the machine to do the predictions on unknown data points, or ask the machine to predict Y for a given X for which you do not know the real value of Y. This training comes under the regression that we talked about earlier.

**Classification:**

You may also use machine learning techniques for classification problems. In classification problems, you classify objects of similar nature into a single group. For example, in a set of 100 students say, you may like to group them into three groups based on their heights – short, medium and long. Measuring the height of each student, you will place them in a proper group. Now, when a new student comes in, you will put him in an appropriate group by measuring his height. By following the principles in regression training, you will train the machine to classify a student based on his feature – the height. When the machine learns how the groups are formed, it will be able to classify any unknown new student correctly. Once again, you would use the test data to verify that the machine has learned your technique of classification before putting the developed model in production. Supervised Learning is where the AI really began its journey. This technique was applied successfully in several cases. You have used this model while doing the hand-written recognition on your machine. Several algorithms have been developed for supervised learning. You will learn about them in the following chapters.

**Unsupervised:**

Learning In unsupervised learning, we do not specify a target variable to the machine, rather we ask machine “What can you tell me about X?”. More specifically, we may ask questions such as given a huge data set X, “What are the five best groups we can make out of X?” or “What features occur together most frequently in X?”. To arrive at the answers to such questions, you can understand that the number of data points that the machine would require to deduce a strategy would be very large. In case of supervised learning, the machine can be trained with even about few thousands of data points. However, in case of unsupervised learning, the number of data points that is reasonably accepted for learning starts in a few millions. These days, the data is generally abundantly available. The data ideally requires curating. However, the amount of data that is continuously flowing in a social area network, in most cases data curation is an impossible task. The unsupervised learning has shown a great success in many modern AI applications, such as face detection, object detection, and so on.

**2.3 ADVANTAGES OF PROPOSED SYSTEM:**

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them. With Machine Learning, you don’t need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

**CHAPTER 3**

**SYSTEM SPECIFICATION**

**3.1 HARDWARE SPECIFICATION:**

Device name: Laptop/Desktop

Processor: INTEL I7 7th gen

Ram: 16GB

Memory: 128 HDD

System Type: 64-bit operating system x64-based processor

**3.2 SOFTWARE SPECIFICATION:**

Platform used: GOOGLE COLLAB

Programming language: PYTHON

Research: GOOGLE

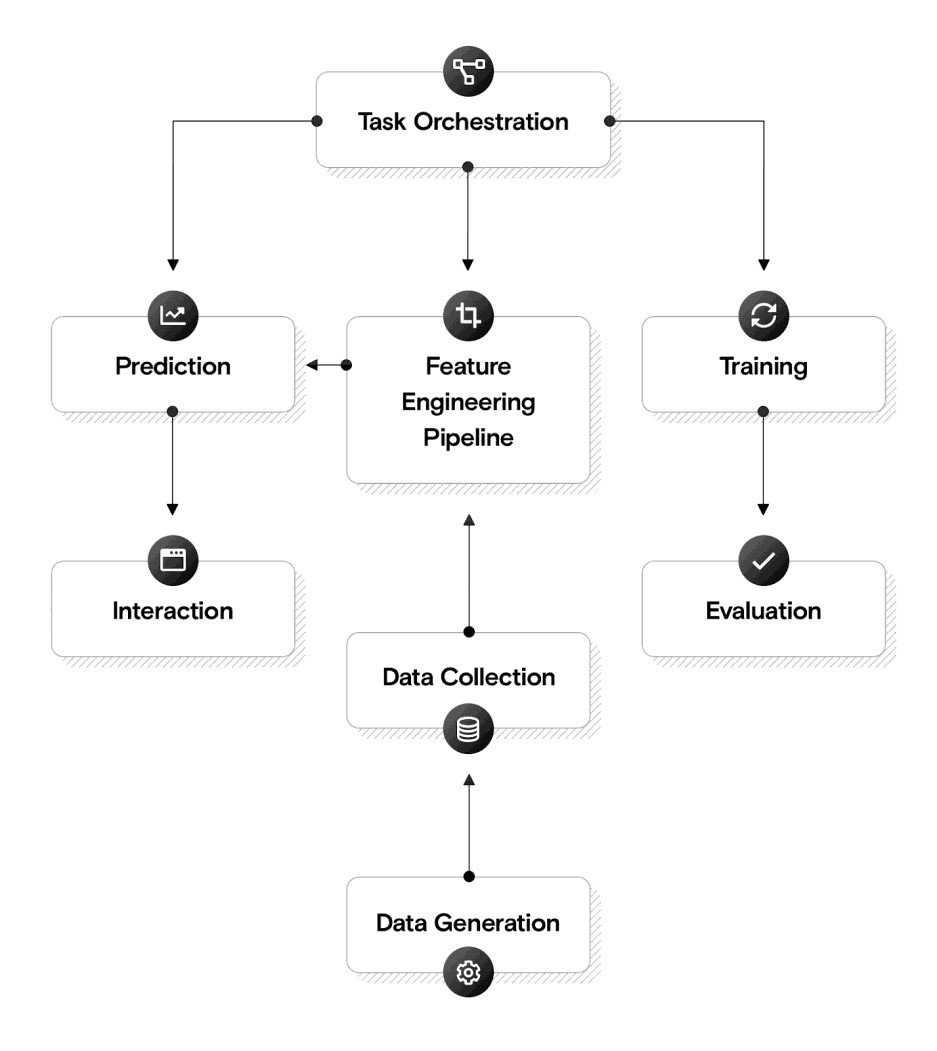
Domain: Machine Learning

**CHAPTER 4**

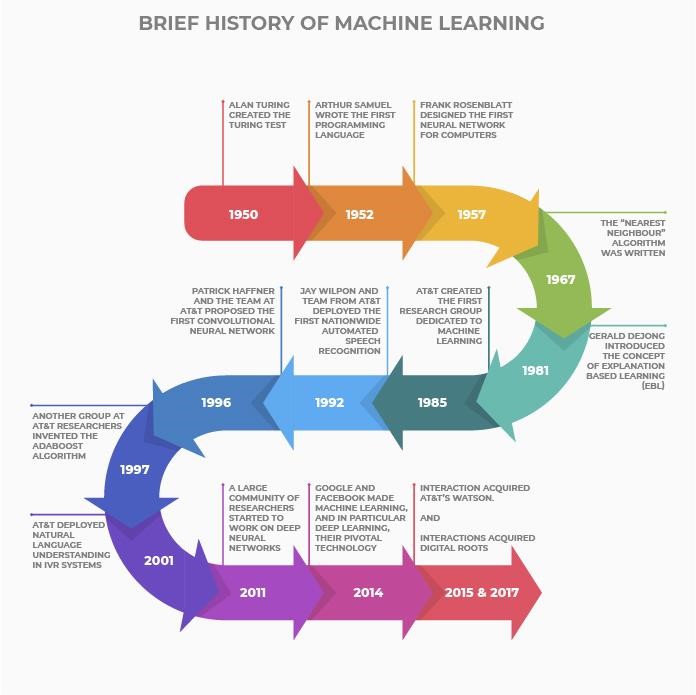
**SYSTEM DESIGN**

**4.1 System Architecture:**

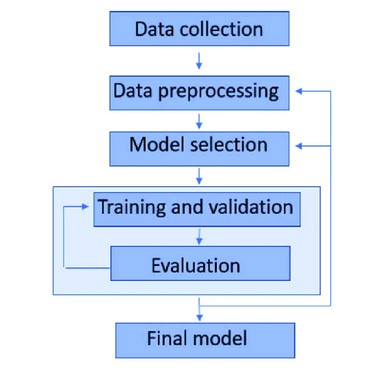
Machine Learning architecture is defined as the subject that has evolved from the concept of fantasy to the proof of reality. As earlier machine learning approach for pattern recognitions has lead foundation for the upcoming major artificial intelligence program. Based upon the different algorithm that is used on the training data machine learning architecture is categorized into three types i.e., Supervised Learning, Unsupervised Learning, and Reinforcement Learning and the process involved in this architecture are Data Acquisition, Data Processing, Model Engineering, Excursion, and Deployment.



**4.2 Module flow Diagrams:**



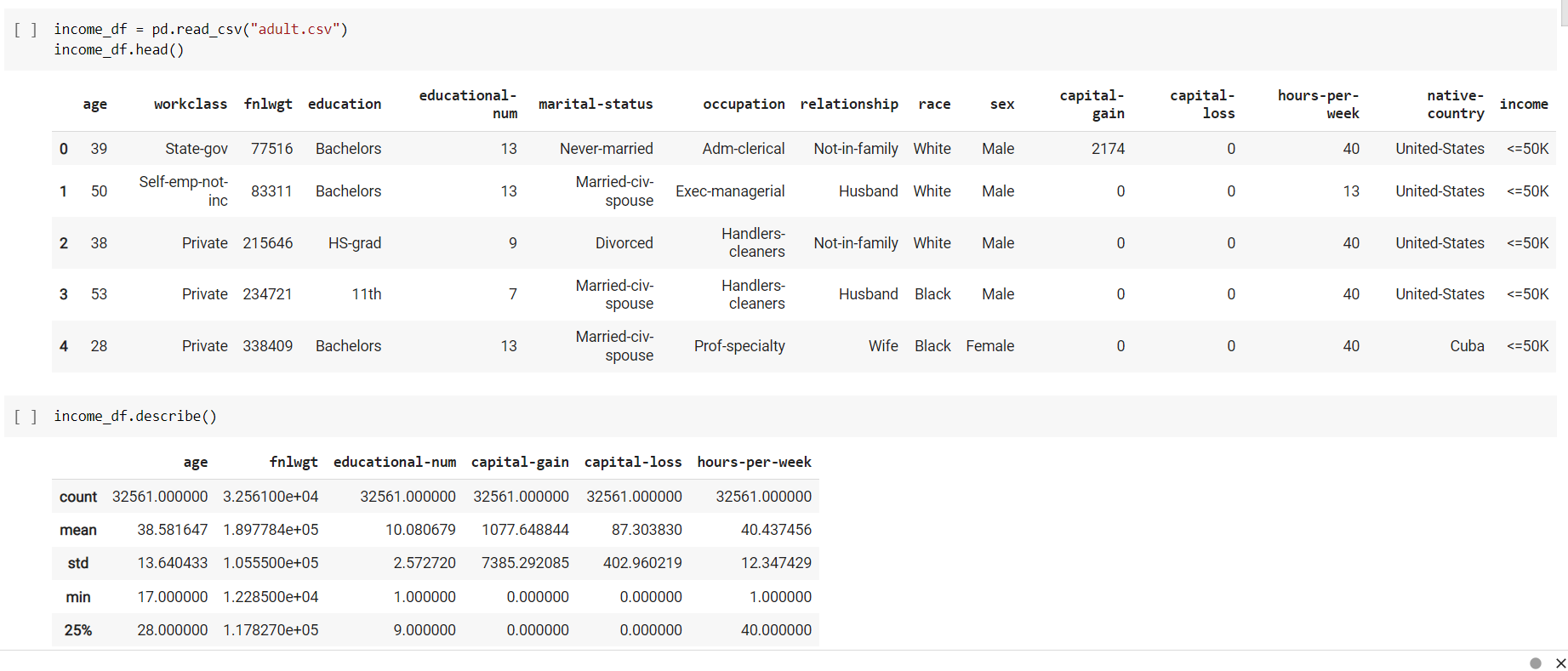
History Of Machine Learning

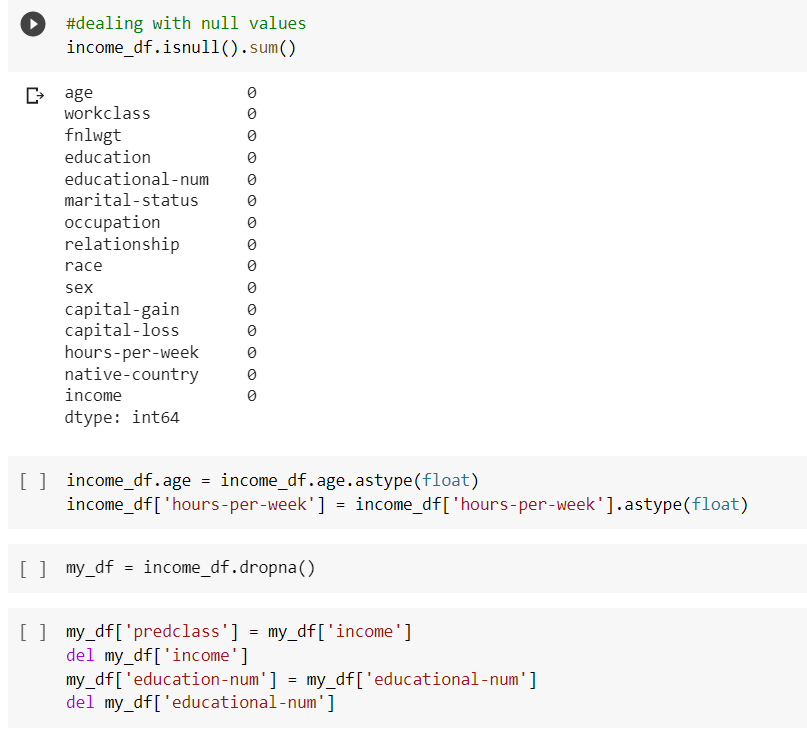


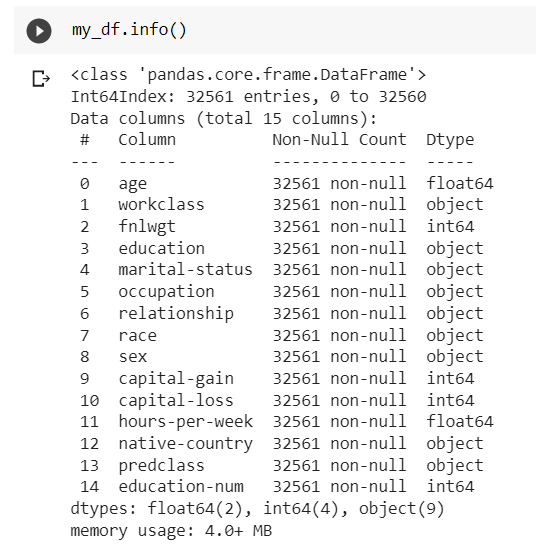
**CHAPTER 5:**

**IMPLEMENTION AND RESULTS**

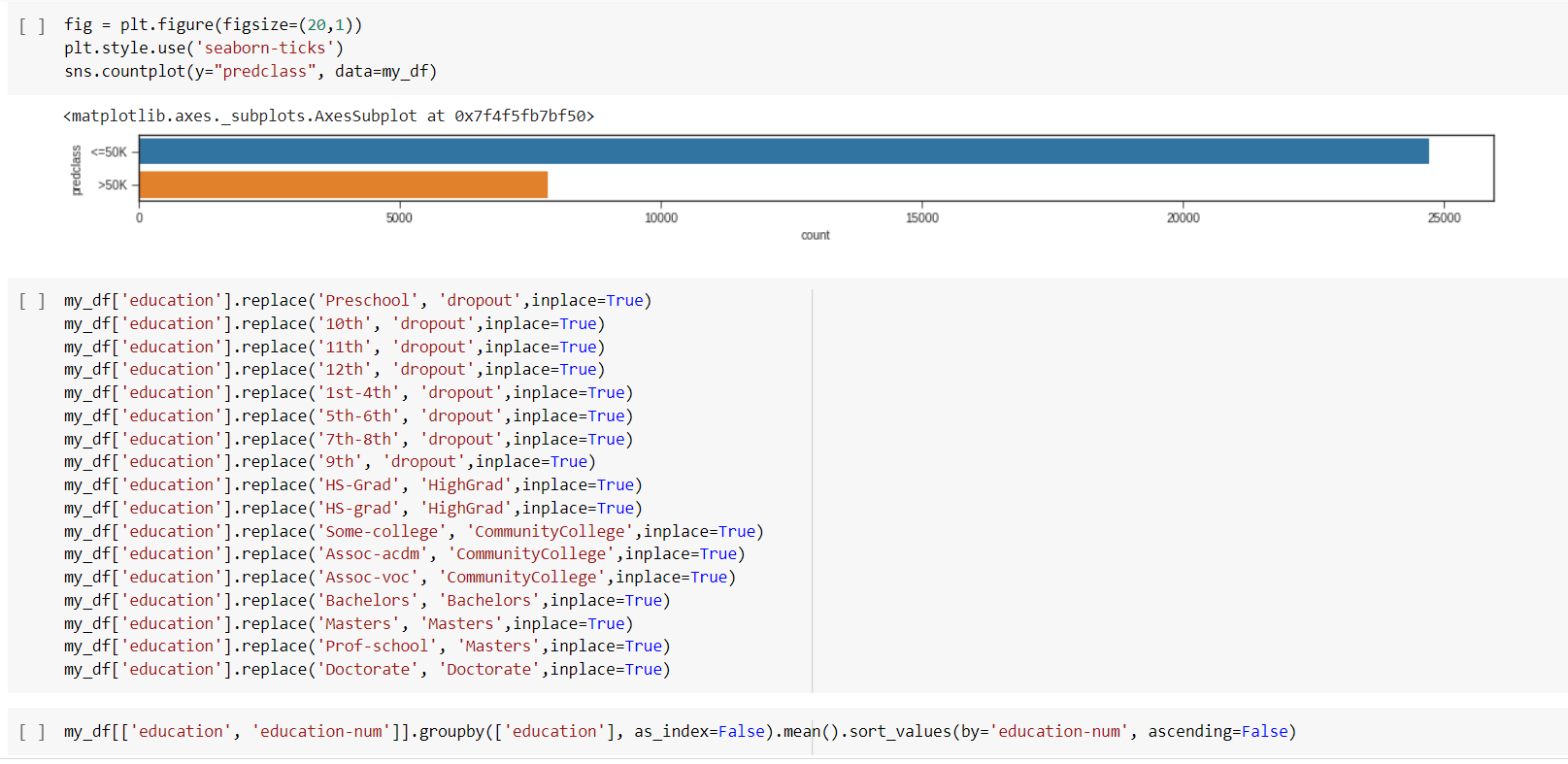
****

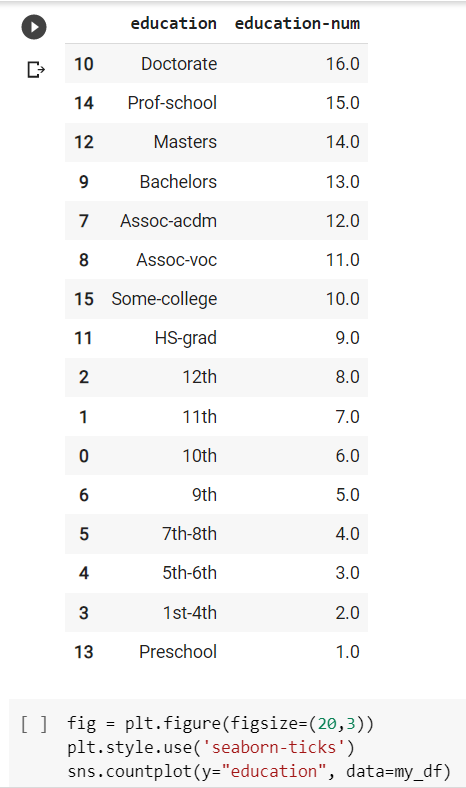
****

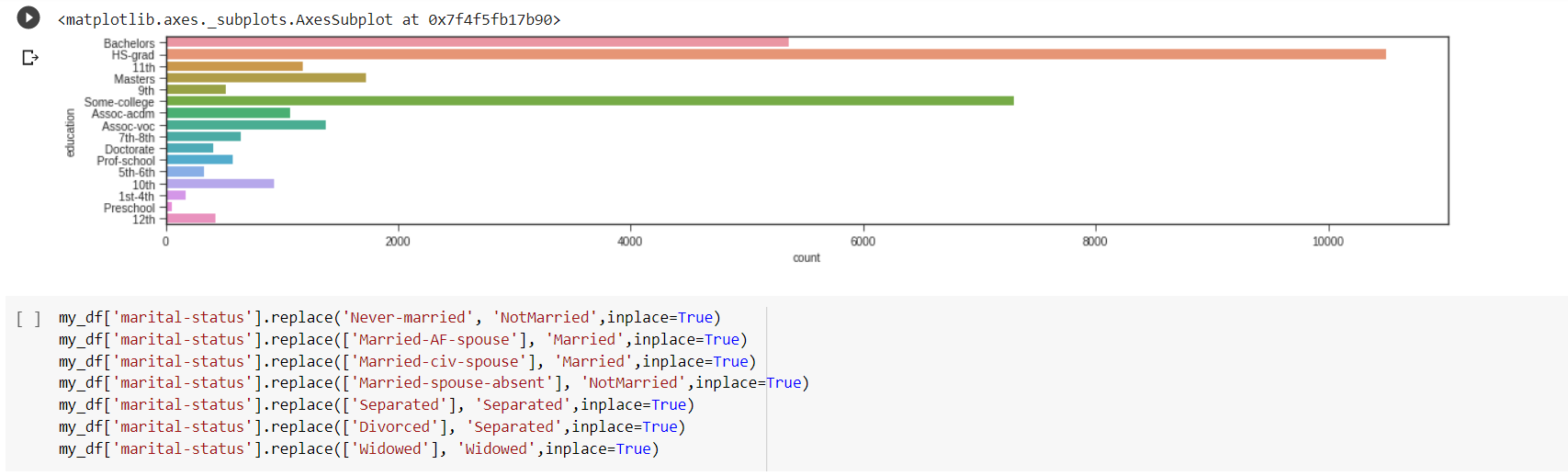
****

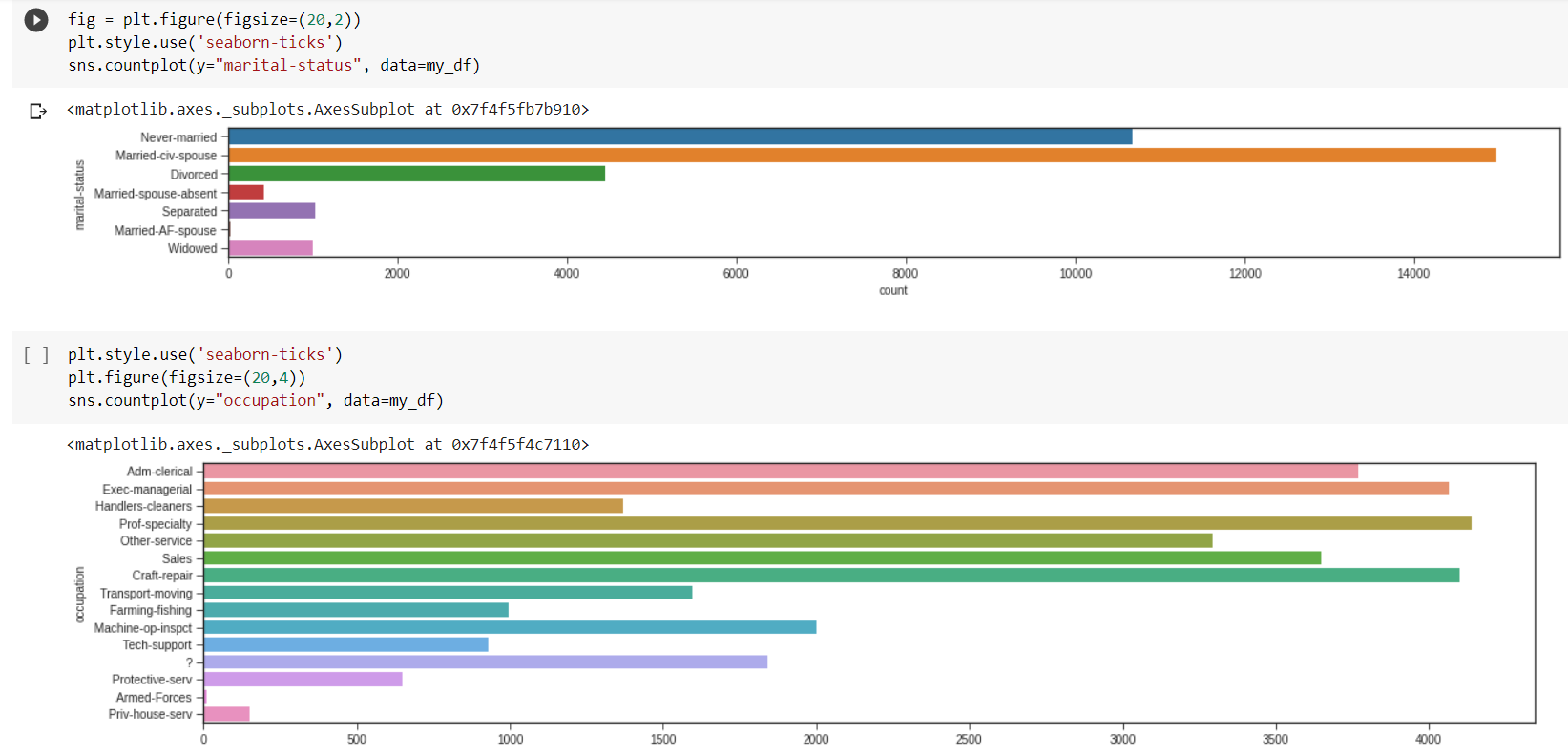
****

****

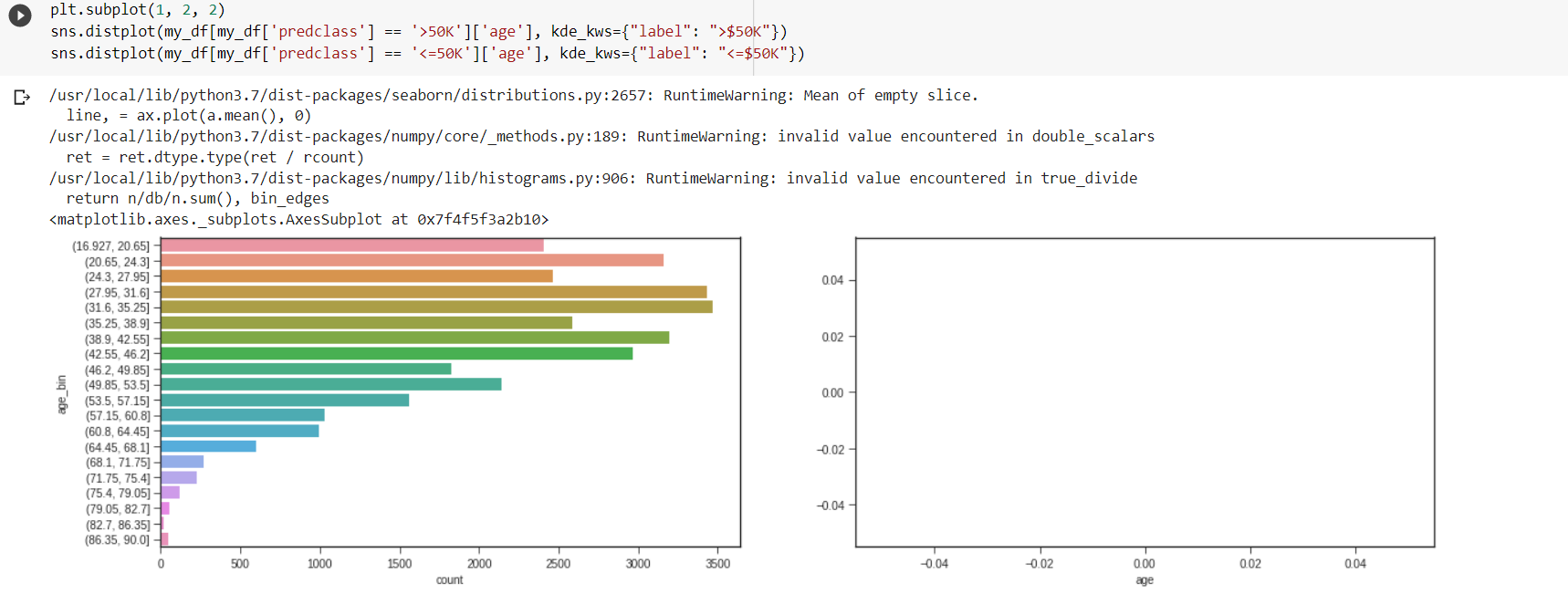
****

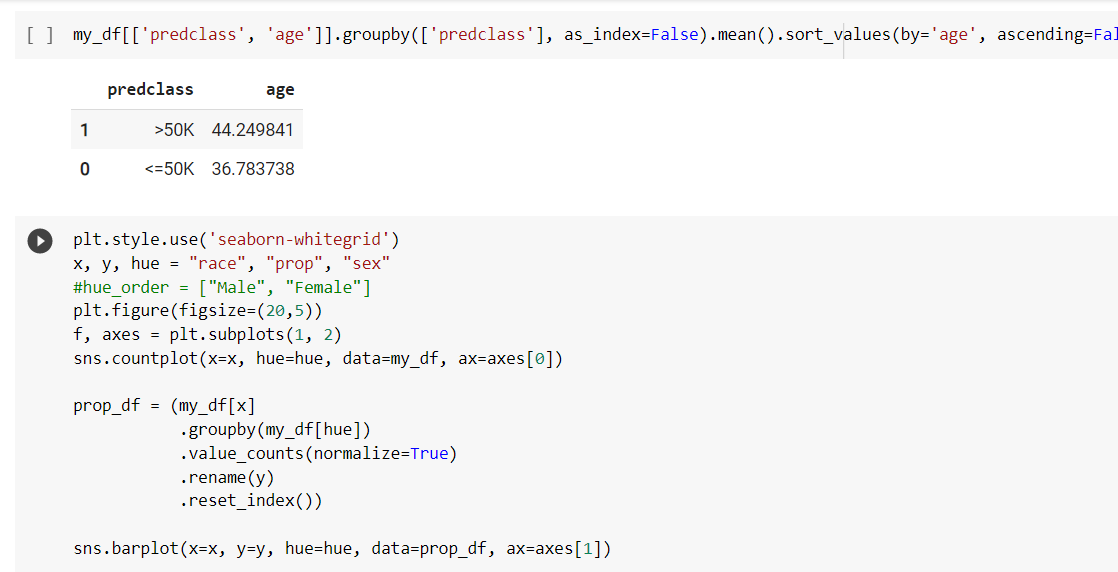
****

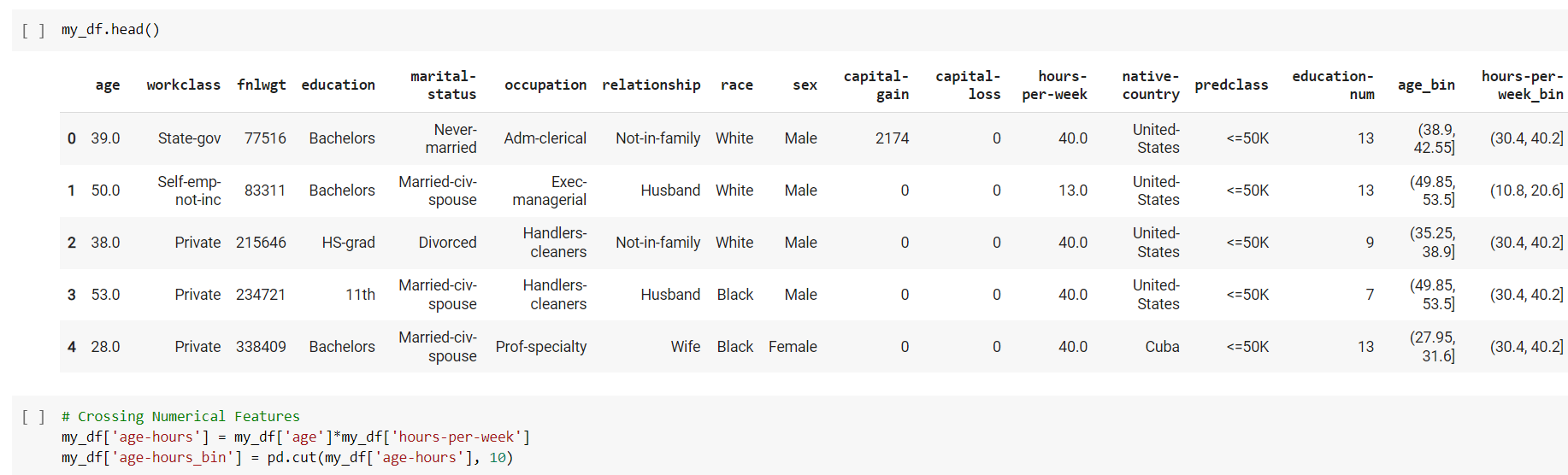
****

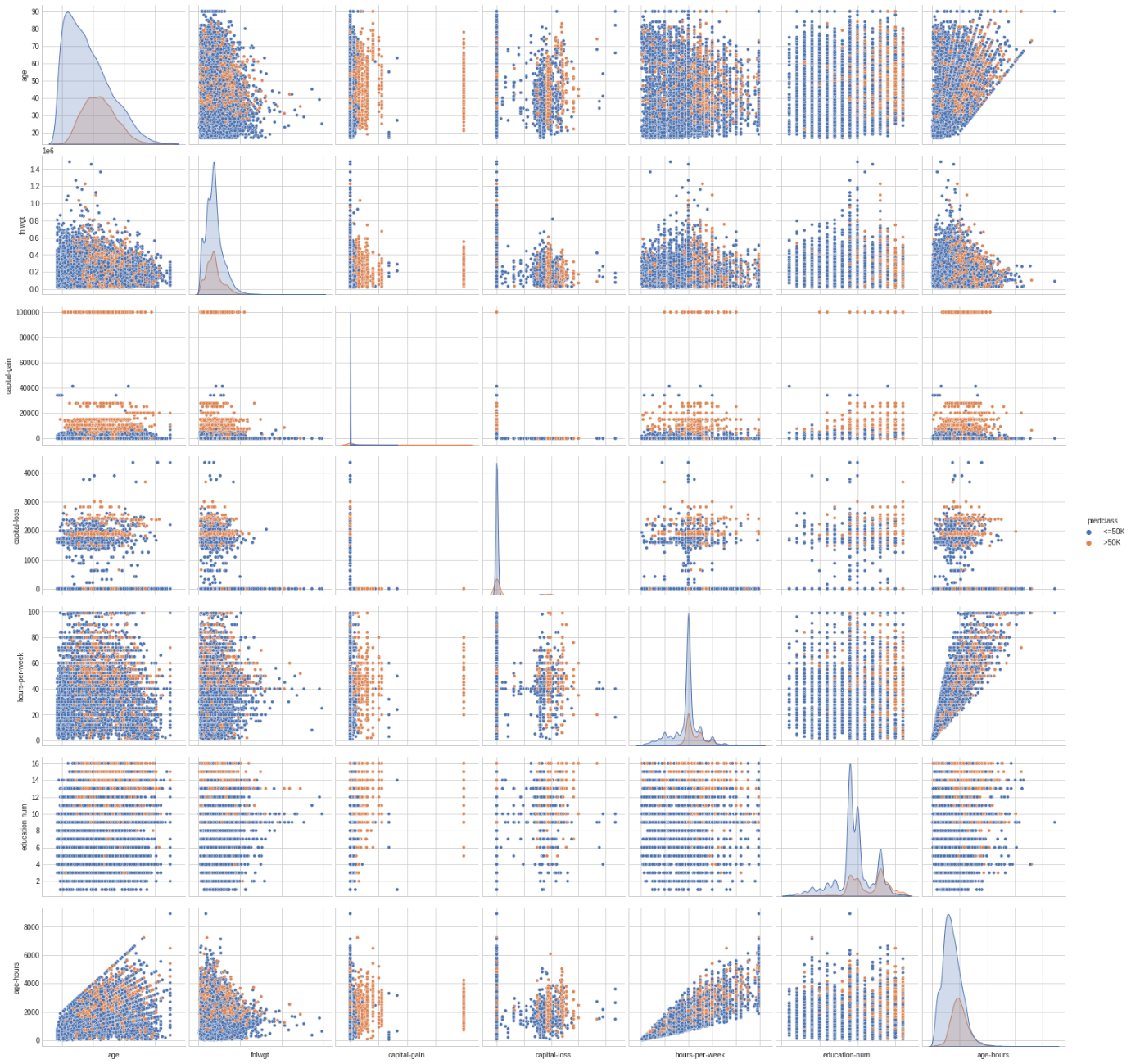
****

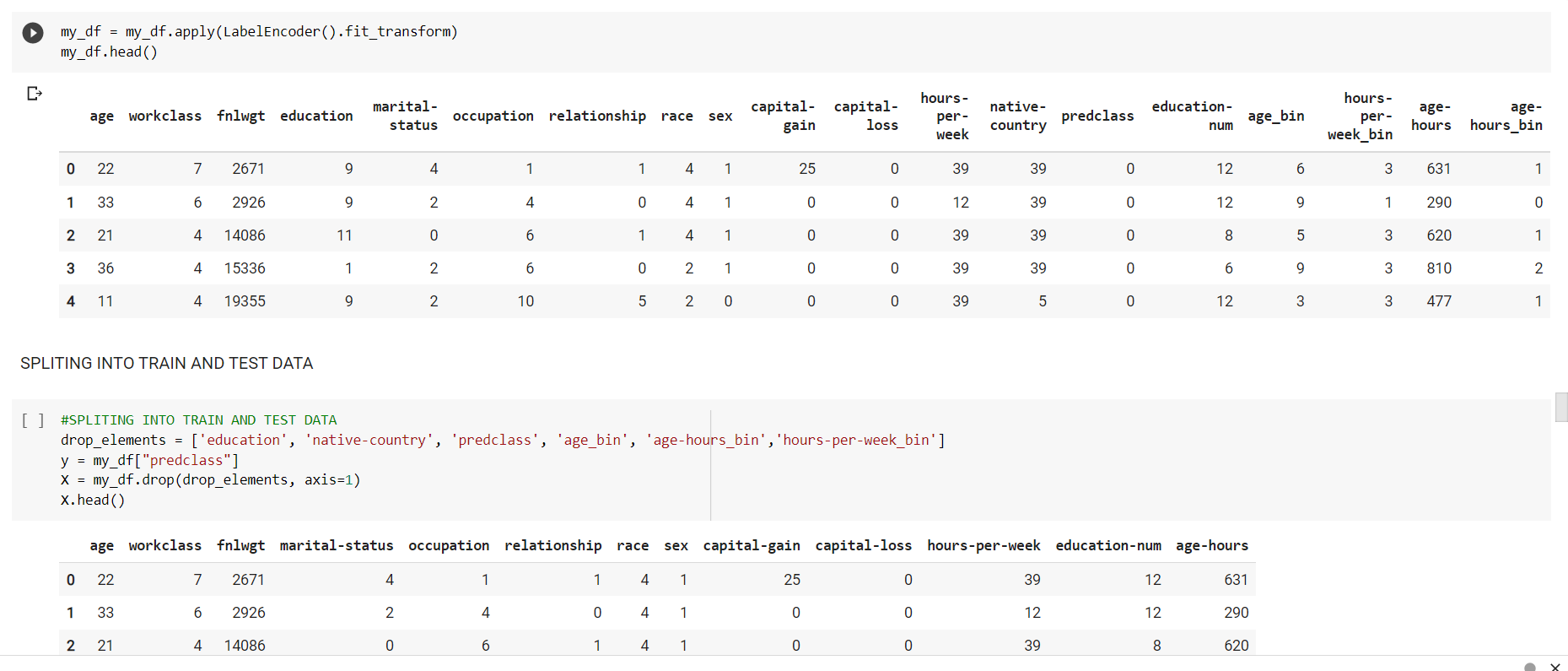
****

****

****

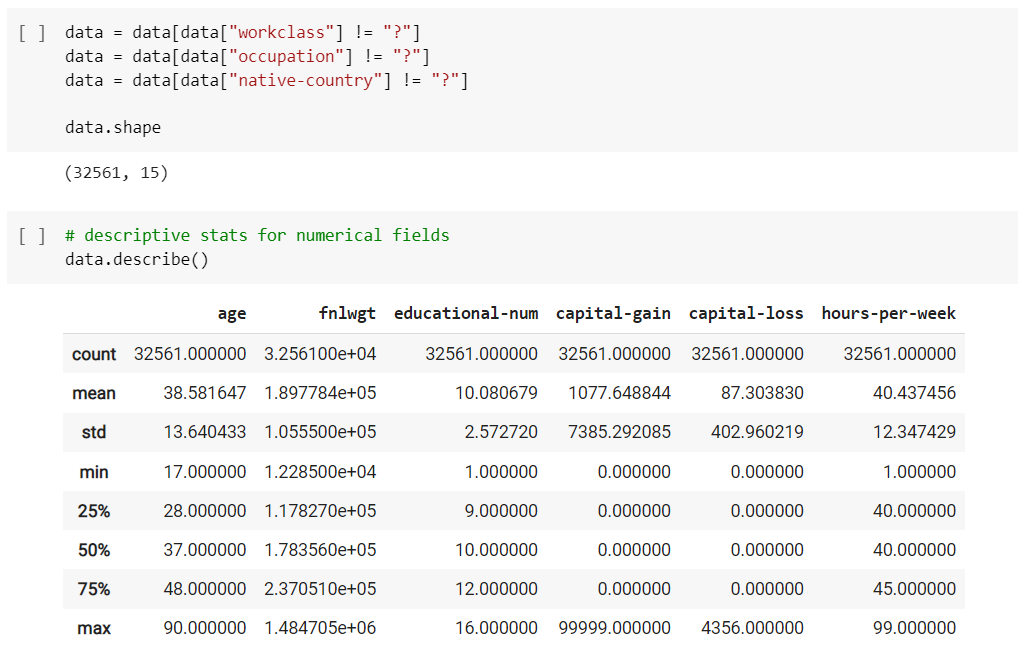
****

****

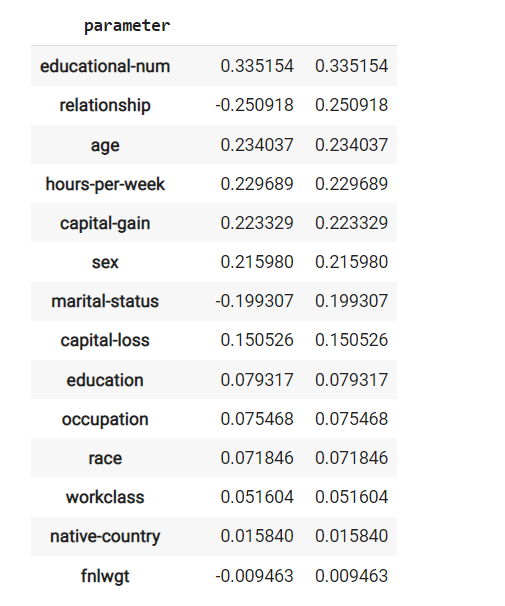
****

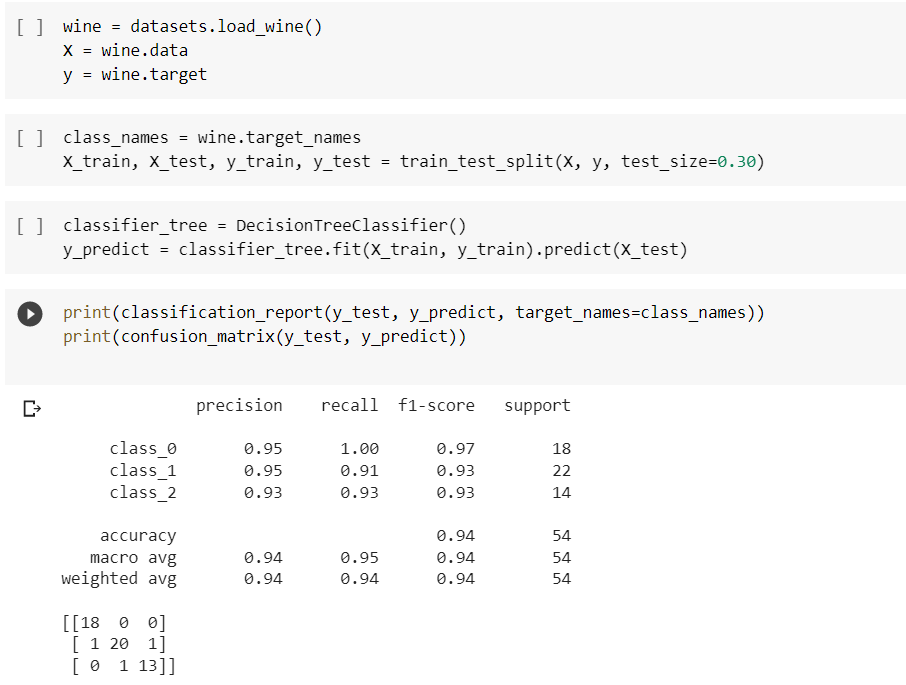
****

****

****

****

****

****

**CHAPTER 6**

**CONCLUSION**

**6.1 CONCLUSION:**

• Overall Internship experience was good, learnt many skills of this particular domain.

• Although it was a virtual internship due to the organization experiences and resources the whole internship was easier.

• This internship is all about learning to me rather than working for projects and satisfying the customer needs.

• However, I think I could have done better in case of developing the projects.

**REFERENCES:**

* Github.com
* Kaggle.com
* Google search
* YouTube