Name: Allan Rodrigues

Class: TE IT A

Roll no: 59

Pid:191104

St. Francis Institute of Technology, Mumbai-400 103 **Department of Information Technology**

A.Y. 2021-2022

Class: TE-ITA/B, Semester: VI Subject: **Business Intelligence Lab**

Experiment – 2: Exercise on Data Exploration

**Aim:** Exercise on Data Exploration.

**Objectives:** After study of this experiment, the student will be able to

* Understand different types of attributes.

**Outcomes:** After study of this experiment, the student will be able to

**CO1:** Understand the importance of data mining along with identification of issues and technologies associated with it.

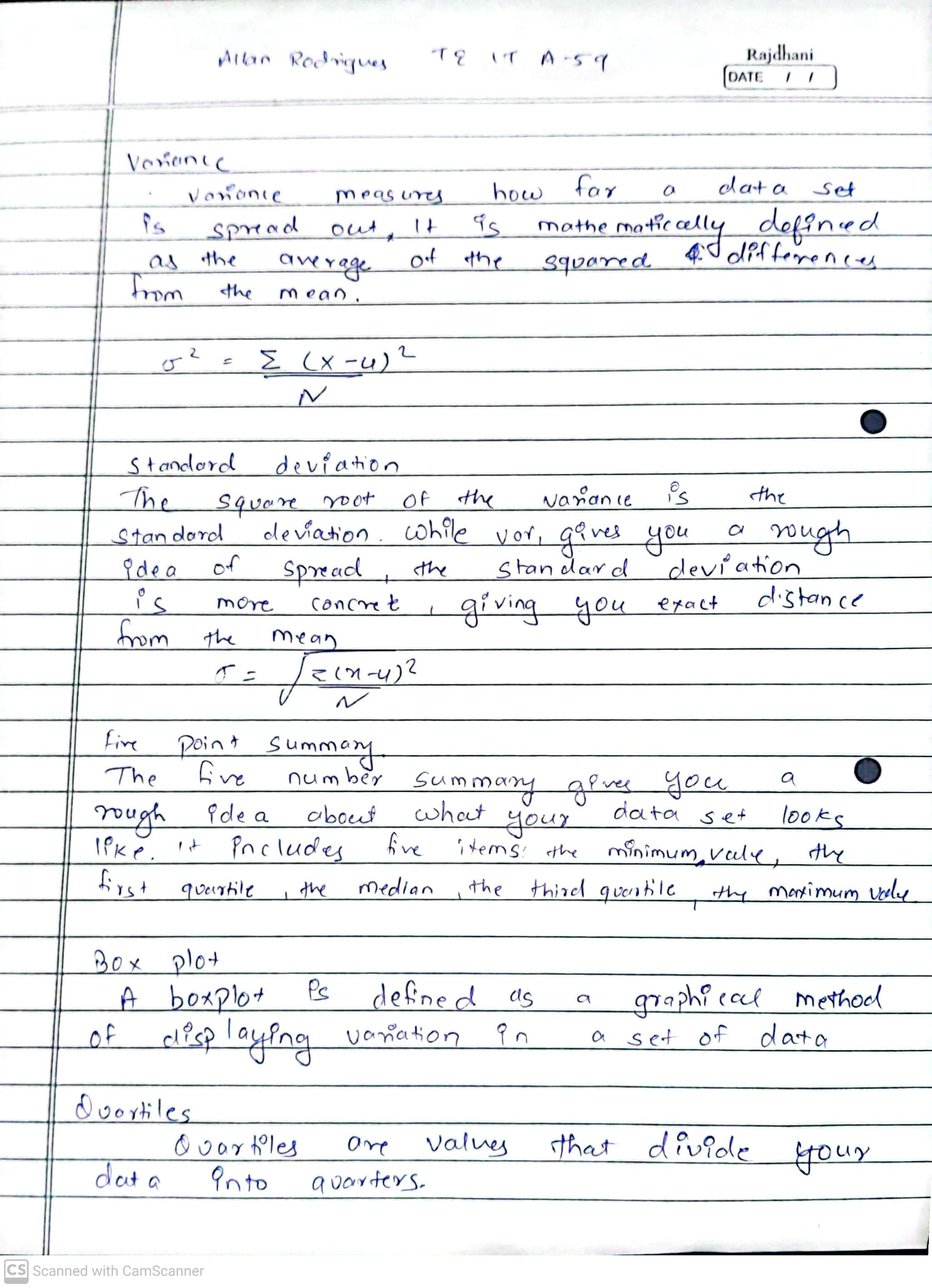
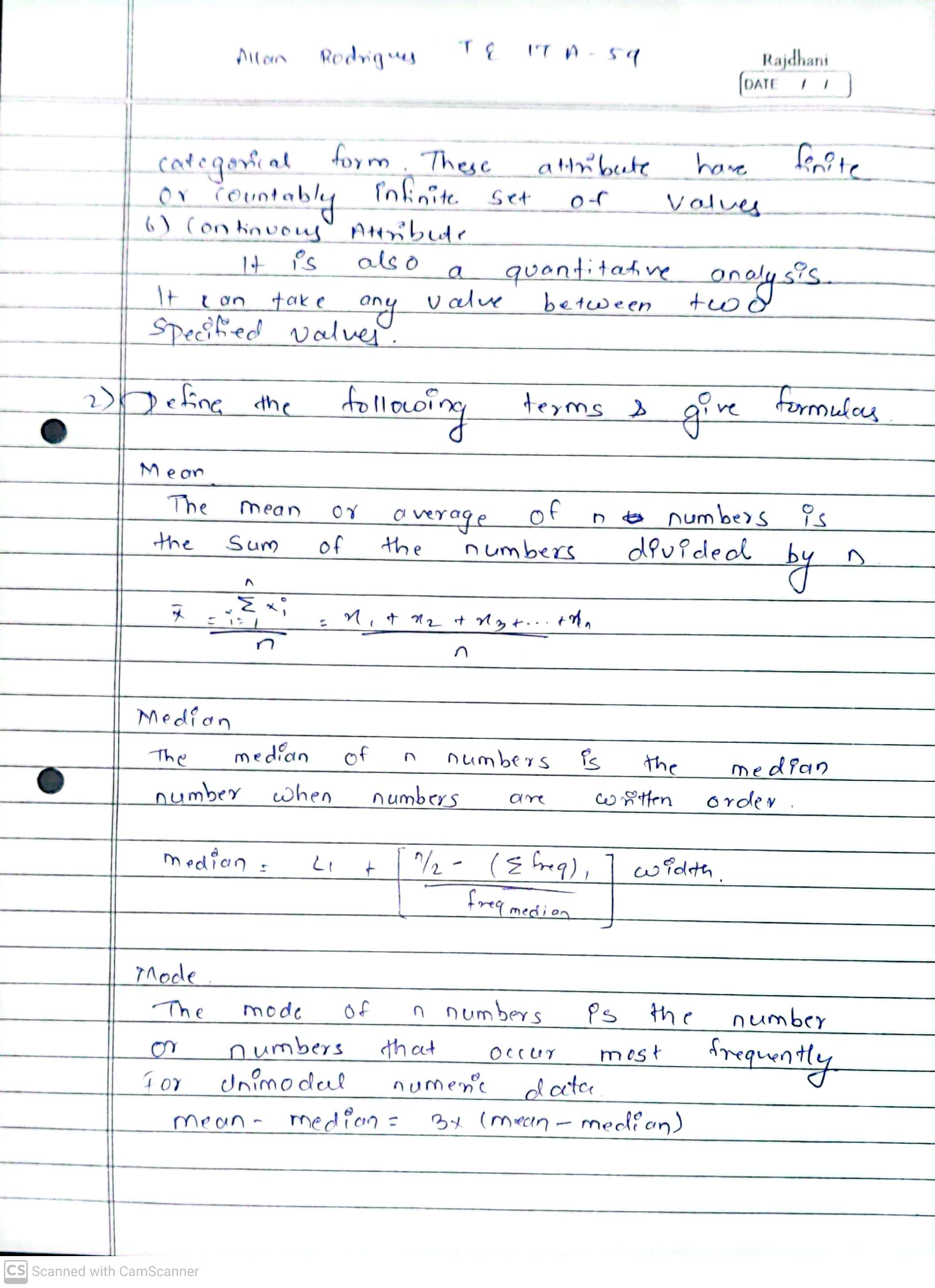
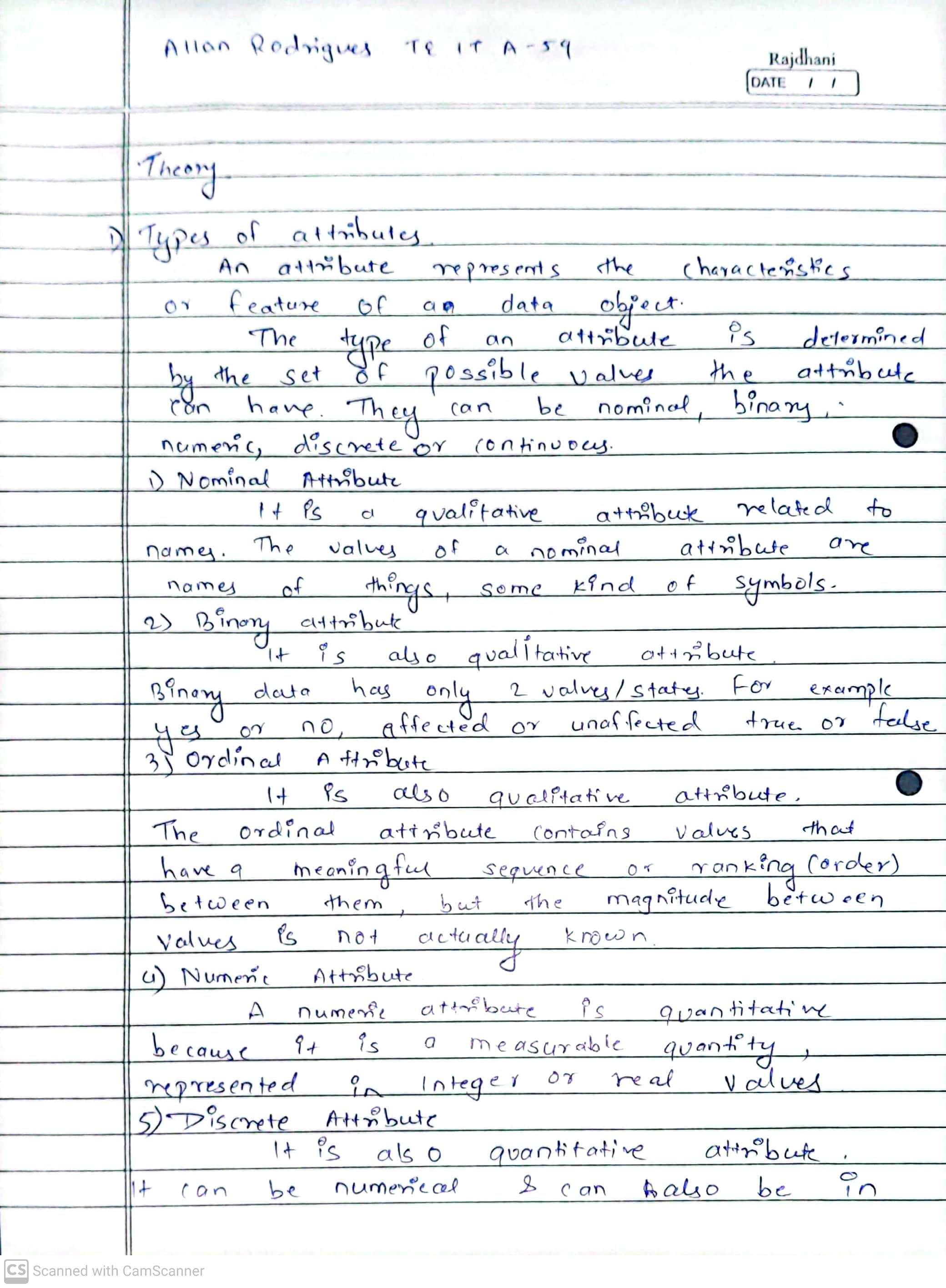
**CO2:** Organize and prepare the data needed for data mining using pre preprocessing techniques and Perform exploratory analysis of the data to be used for mining

**Prerequisite:** Introduction to different types of attributes.

**Requirements:** Personal Computer, Windows XP operating system, Internet Connection, Microsoft Word, WEKA tool.

# Theory: (hand written)

* Types of attributes
* Define the following terms and give formulas
  + Mean, Median, Mode, Variance, Standard deviation, Five number summary, Box plot, Range, Quartile, Interquartile range



* Solve following problems:
  1. Suppose that value for given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows:

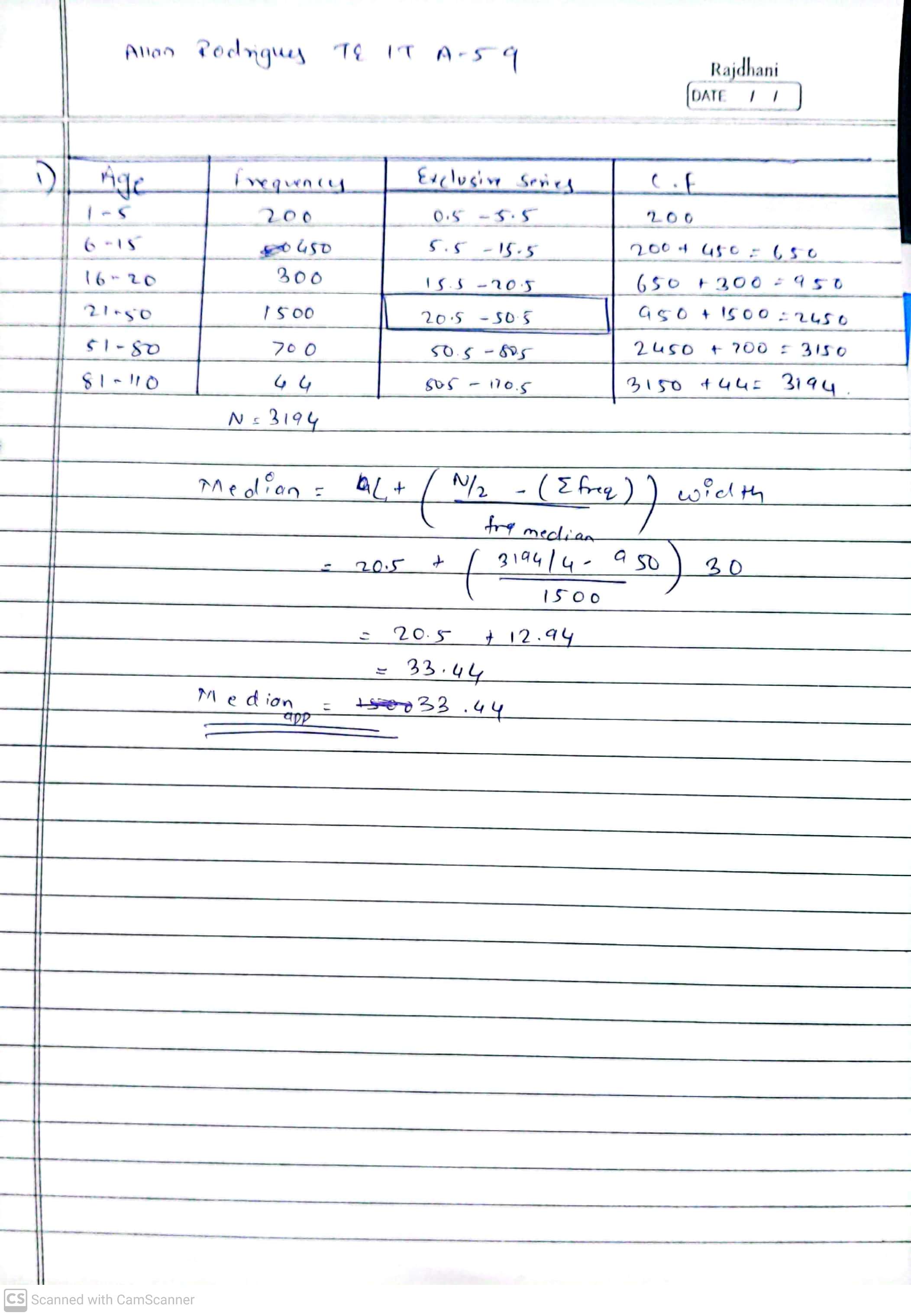
|  |  |
| --- | --- |
| Age | Frequency |
| 1-5 | 200 |
| 6-15 | 450 |
| 16-20 | 300 |
| 21-50 | 1500 |
| 51-80 | 700 |

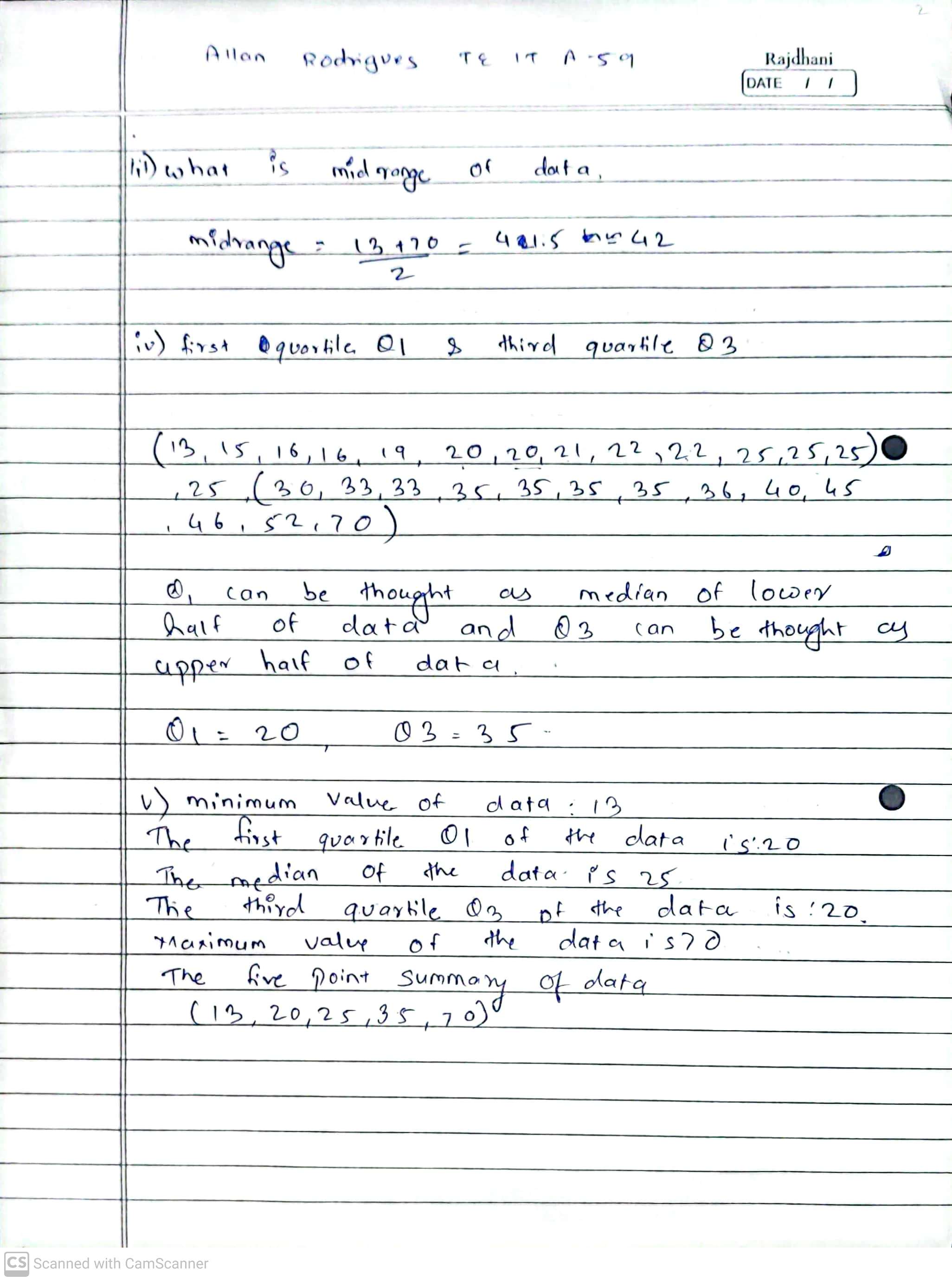
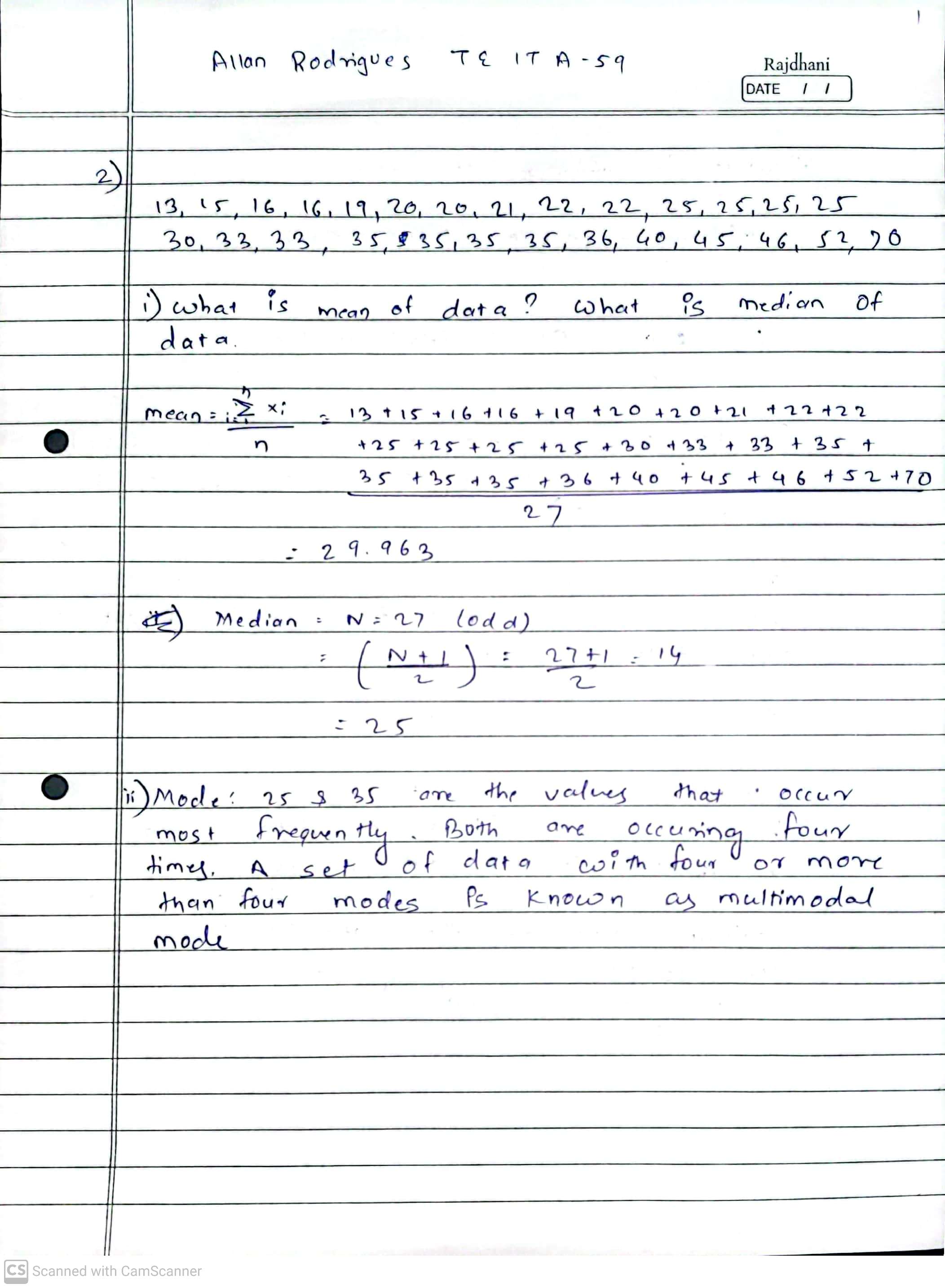
44

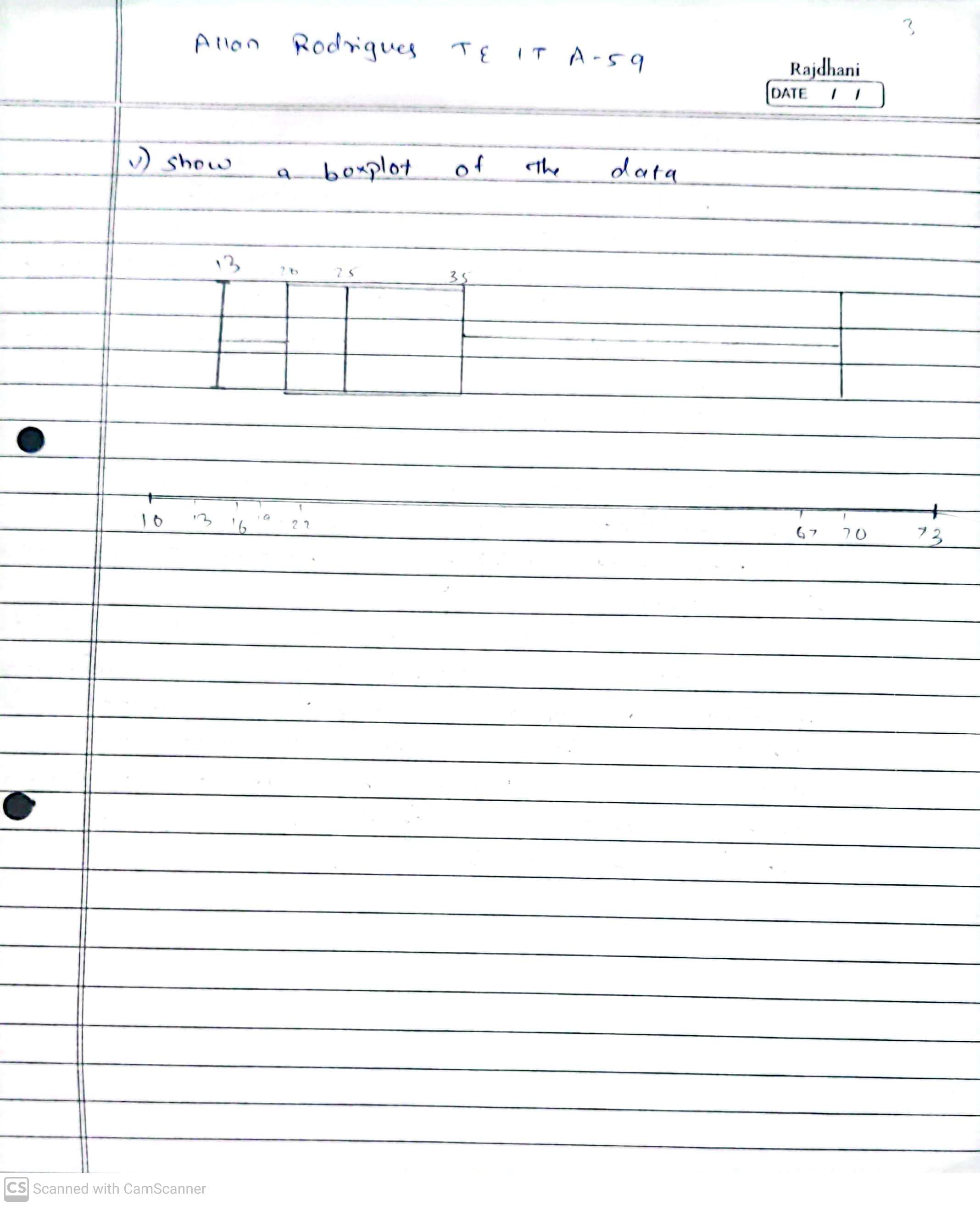
81-110

Compute an approx. median value for the data.

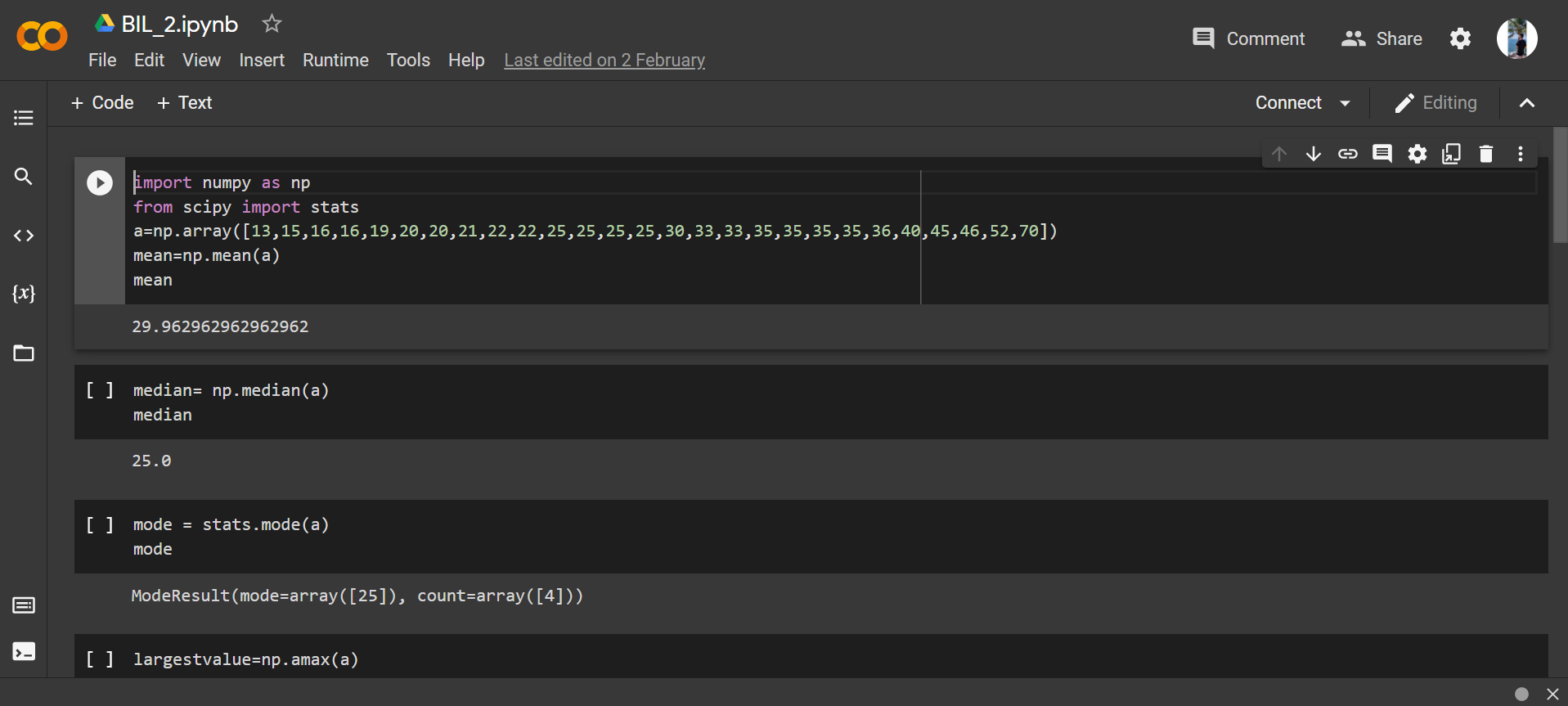
* 1. Suppose that data for analysis includes the attribute age. The age values for data tuples are(in increasing order): 13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70
     1. What is mean of data? What is median of data?
     2. What is mode of data? Comment on data's modality (bimodal/trimodal etc.)
     3. What is mid range of data?(smallest+ largest value/2)
     4. Can you find roughly the first quartile Q1, and the third quartile Q3 of the data.
     5. Give the five point summary of the data.
     6. Show a box plot of the data.

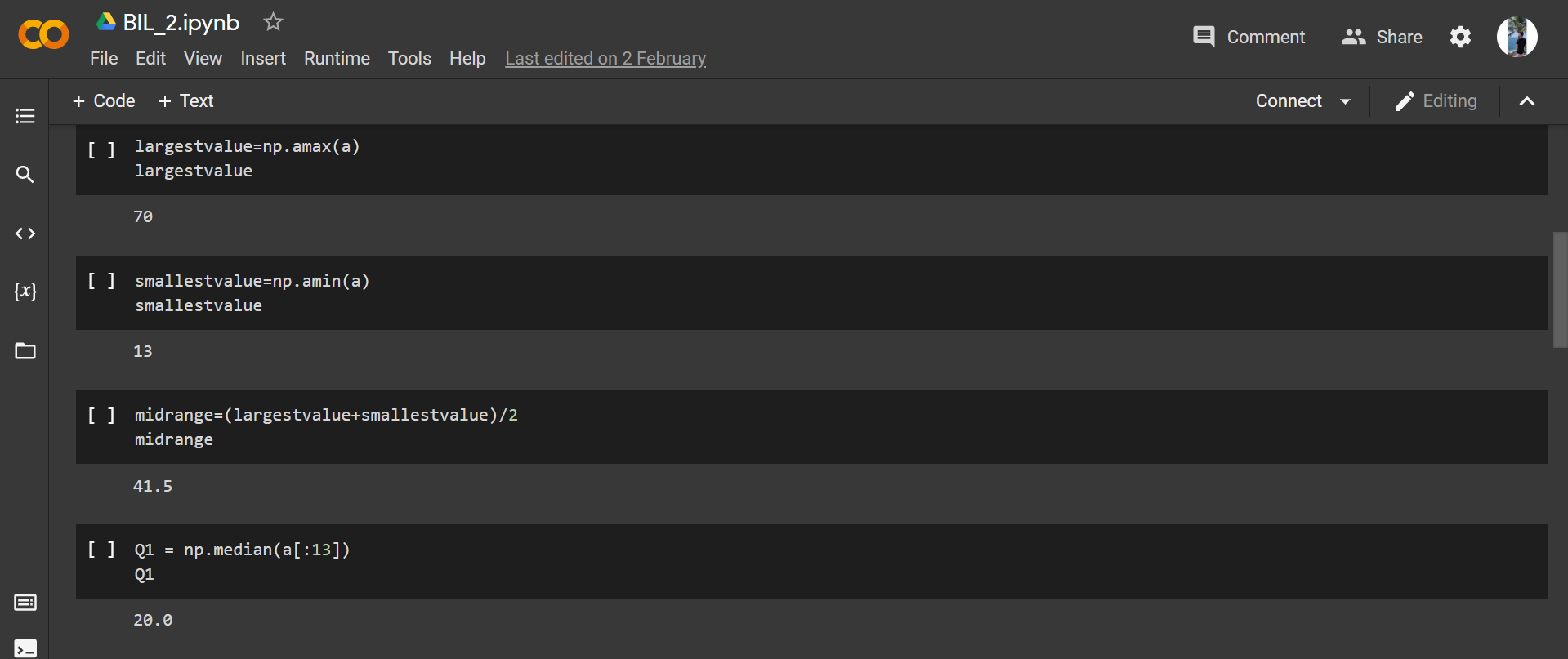


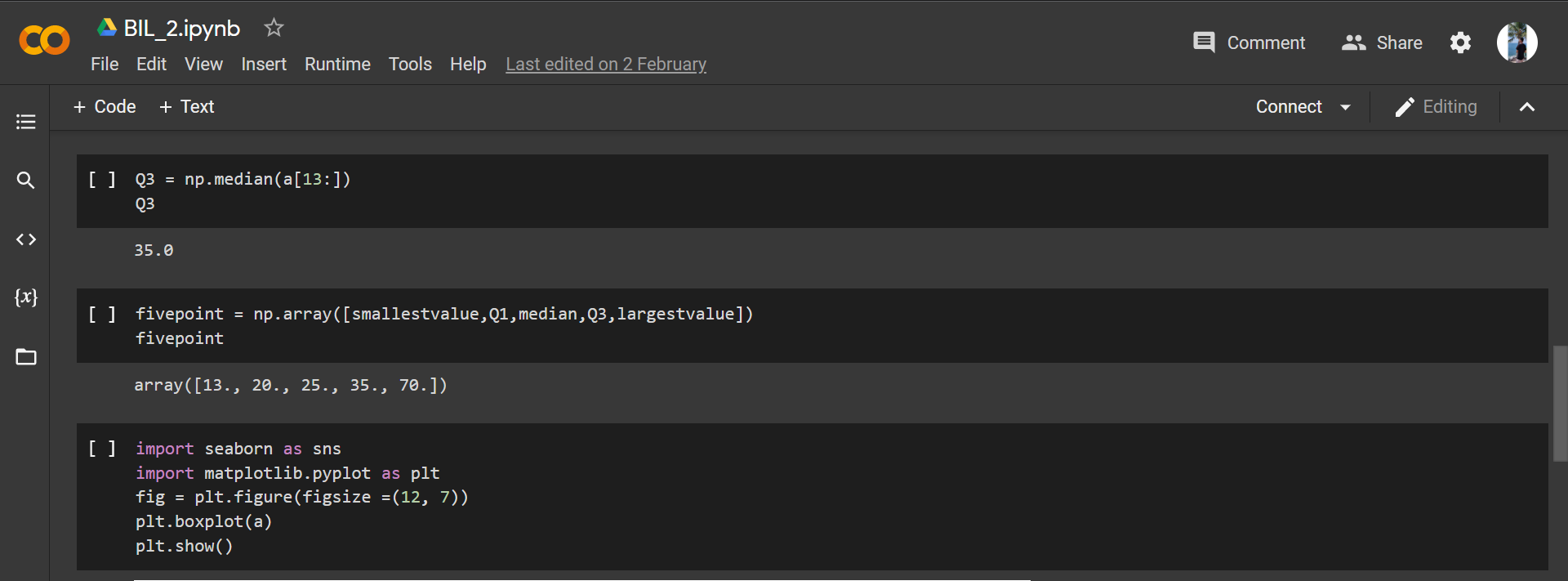


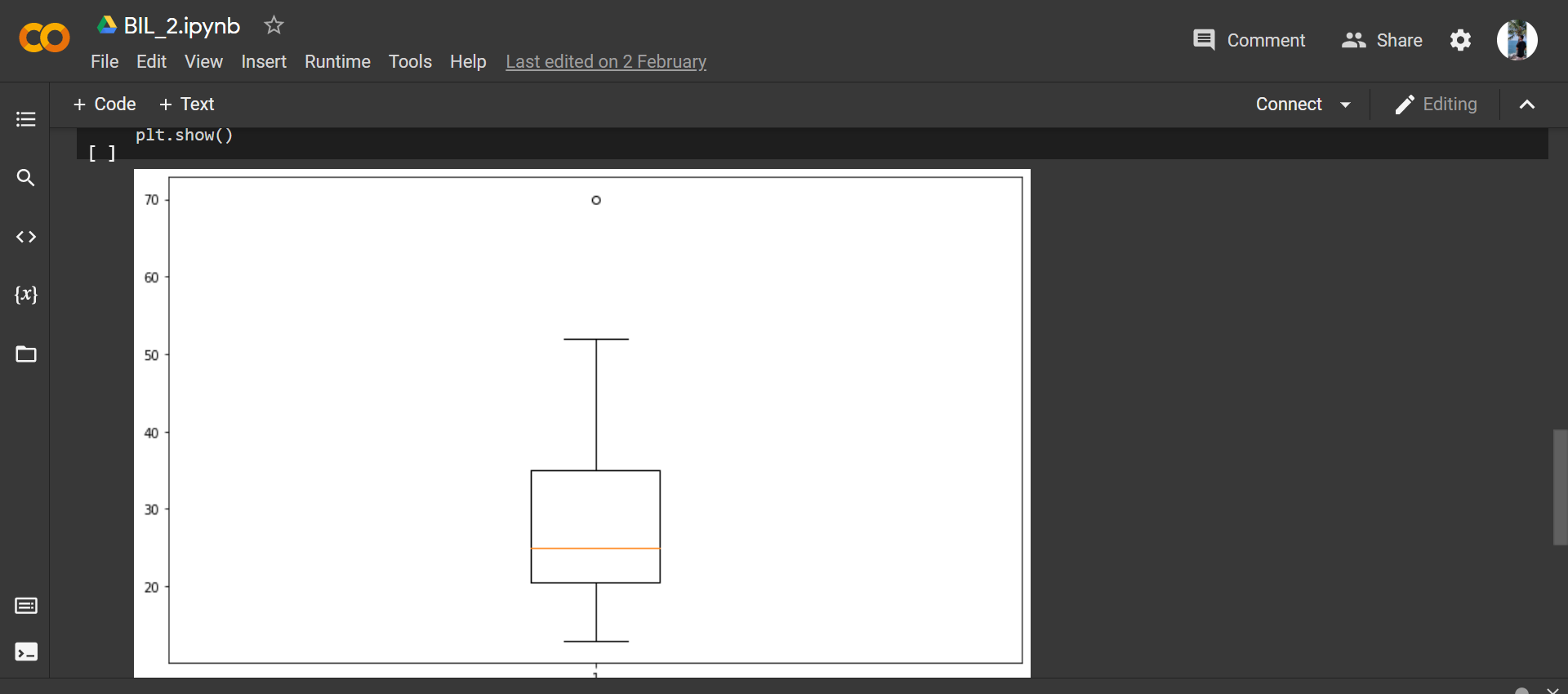


**Laboratory Exercise: Write Java / Python code for ablove problem and attach** Printout of code along with output Snapshots



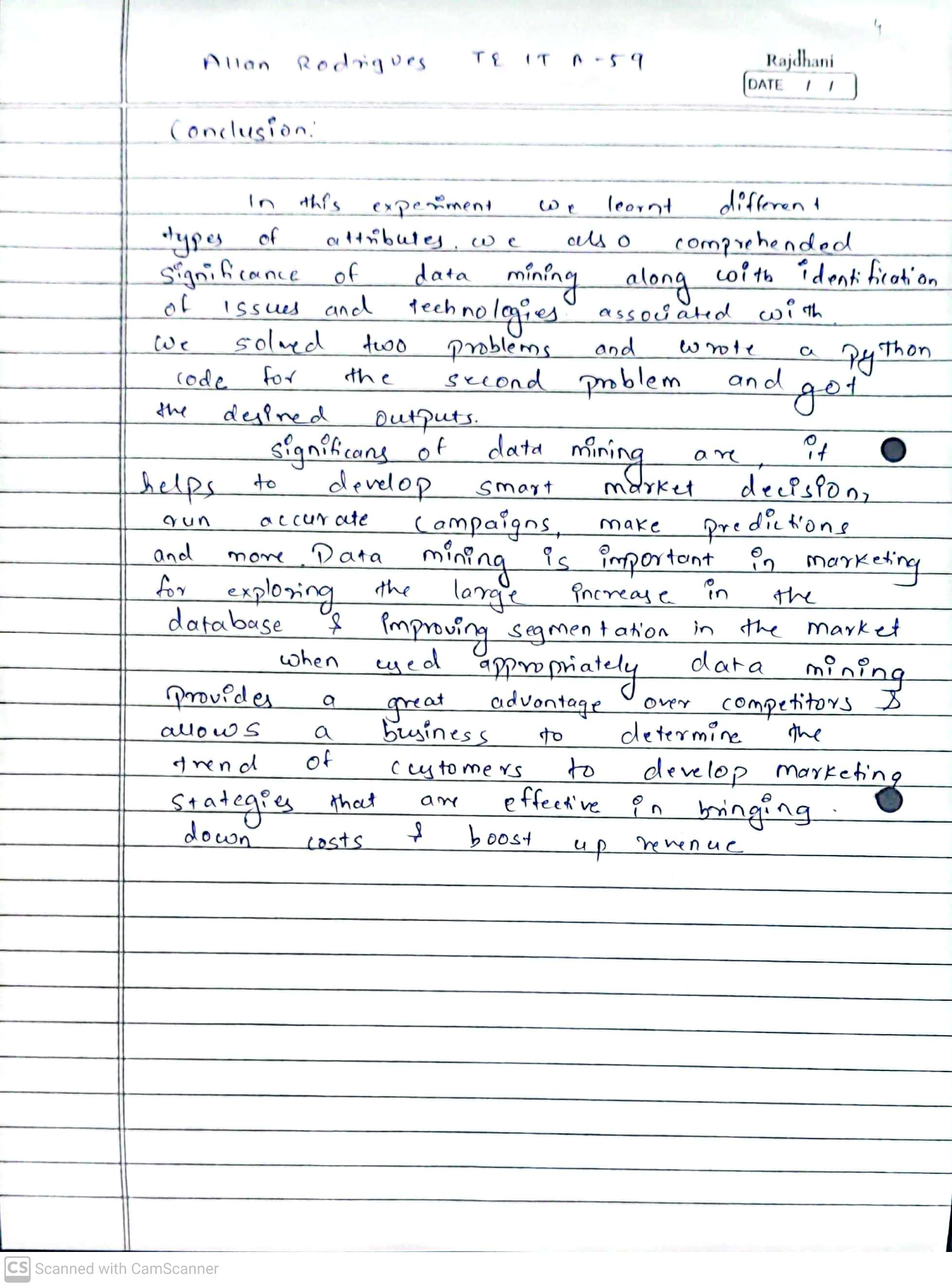






# Post-Experiments Exercise

1. **Conclusion:**
   1. Summary of Experiment
   2. Importance of Experiment
   3. Application of Experiment



**Reference:** Data Mining: Concept & Techniques, 3rd Edition, Jiawei Han, Micheline Kamber, Jian Pei, Elsevier.