**Digital Assessment 1**

**CBS3014 – Modern Web Applications**

Date: 15 August, 2024

Name: Anuj Parihar

Registration Number: 21BBS0162

**Link to Assessment Codebase and Dataset**:

<https://github.com/BearTS/data-mining-assignments/tree/main/Lab/DA%201>

1. Create Employee Dataset in excel and store the file in .csv extension

|  |  |  |
| --- | --- | --- |
| Attributes | Type | Range of Values |
| Empid | Nominal | 1,2….30 |
| Name | Character | … |
| Designation | Ordinal | Manager, Supervisor, Clerk, labour |
| Salary | Numerical | 6000,….30,0000 |
| Experience | Numerical | 1,2…..12 |
| Vaccinated | Binary | Yes,no |

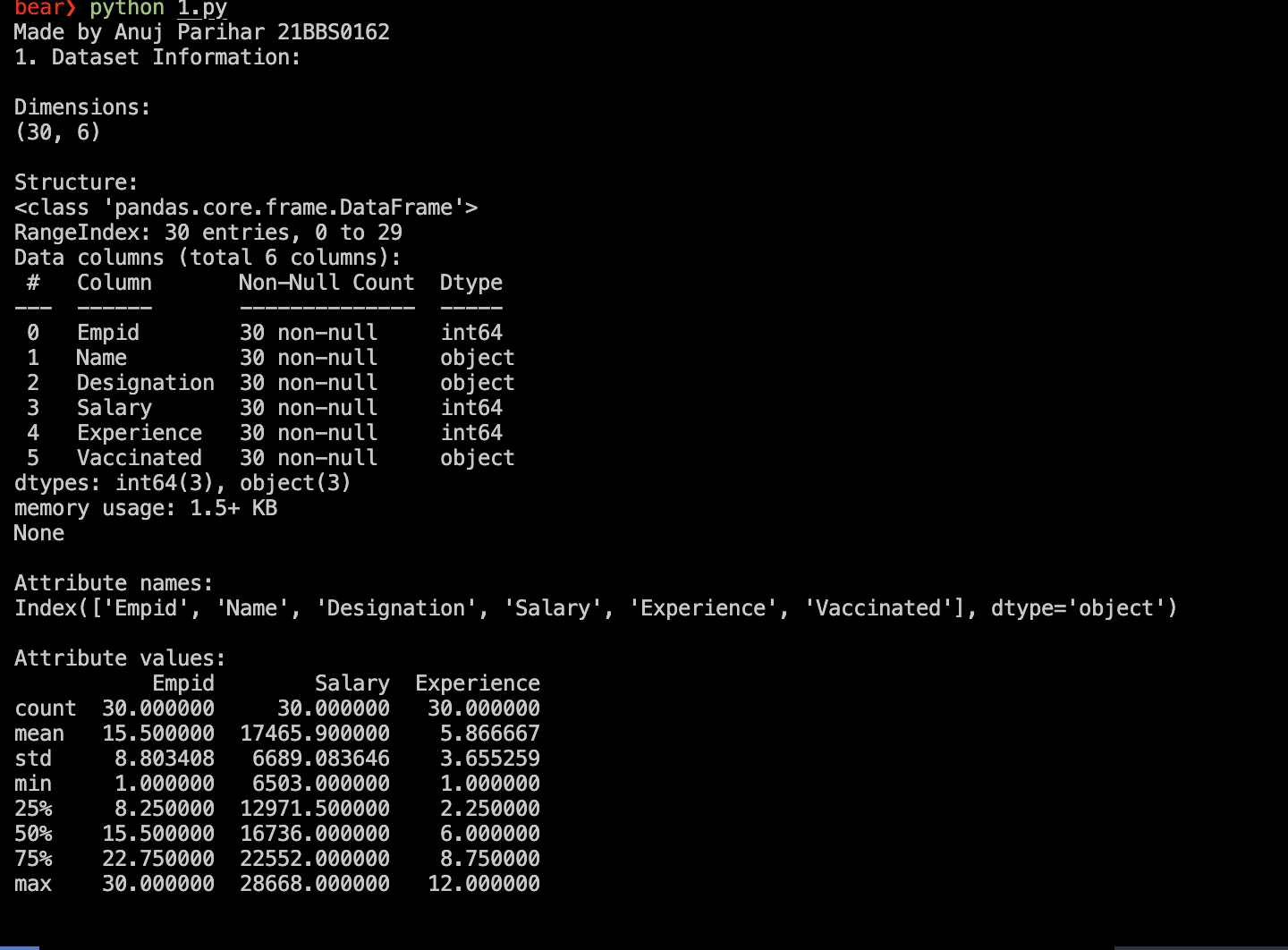
For this question I am using this table data:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Empid | Name | Designation | Salary | Experience | Vaccinated |
| 1 | John Smith | Labour | 27542 | 11 | no |
| 2 | Jane Doe | Clerk | 20637 | 1 | no |
| 3 | Robert Brown | Supervisor | 6750 | 11 | yes |
| 4 | Emily Davis | Manager | 22135 | 7 | no |
| 5 | Michael Wilson | Supervisor | 17295 | 4 | yes |
| 6 | Mary Johnson | Labour | 18535 | 10 | no |
| 7 | David White | Labour | 19837 | 10 | no |
| 8 | Jessica Taylor | Labour | 22093 | 1 | no |
| 9 | Daniel Anderson | Clerk | 14186 | 9 | no |
| 10 | Emma Thomas | Supervisor | 10334 | 8 | no |
| 11 | Matthew Moore | Manager | 23811 | 6 | no |
| 12 | Amanda Jackson | Labour | 14991 | 1 | no |
| 13 | Joshua Harris | Supervisor | 8131 | 1 | yes |
| 14 | Ashley Martin | Clerk | 26156 | 8 | no |
| 15 | Christopher Clark | Manager | 22691 | 2 | yes |
| 16 | Sarah Lewis | Clerk | 6503 | 6 | yes |
| 17 | Andrew Walker | Manager | 26870 | 8 | no |
| 18 | Elizabeth Young | Labour | 27396 | 1 | yes |
| 19 | Joseph Hall | Supervisor | 13225 | 5 | no |
| 20 | Patricia Allen | Supervisor | 15058 | 1 | yes |
| 21 | Thomas King | Manager | 12887 | 8 | no |
| 22 | Linda Scott | Labour | 6975 | 4 | yes |
| 23 | Mark Green | Manager | 14012 | 10 | no |
| 24 | Barbara Adams | Manager | 11464 | 10 | no |
| 25 | Paul Baker | Supervisor | 24958 | 1 | no |
| 26 | Susan Miller | Clerk | 12724 | 5 | yes |
| 27 | Steven Turner | Clerk | 28668 | 3 | no |
| 28 | Karen Nelson | Supervisor | 16887 | 5 | no |
| 29 | Charles Carter | Labour | 14641 | 7 | no |
| 30 | Nancy Mitchell | Manager | 16585 | 12 | yes |

**Aim:** Get the dimensions, structure, attribute name, and attribute values of the dataset

**Input:** The input dataset is mentioned above

**Output:**

****

**Results:**In this experiment we get to know the attribute names and different types of data in the csv

**Aim:** Display:   
 1. First 5 Records

2. Last 5 Records

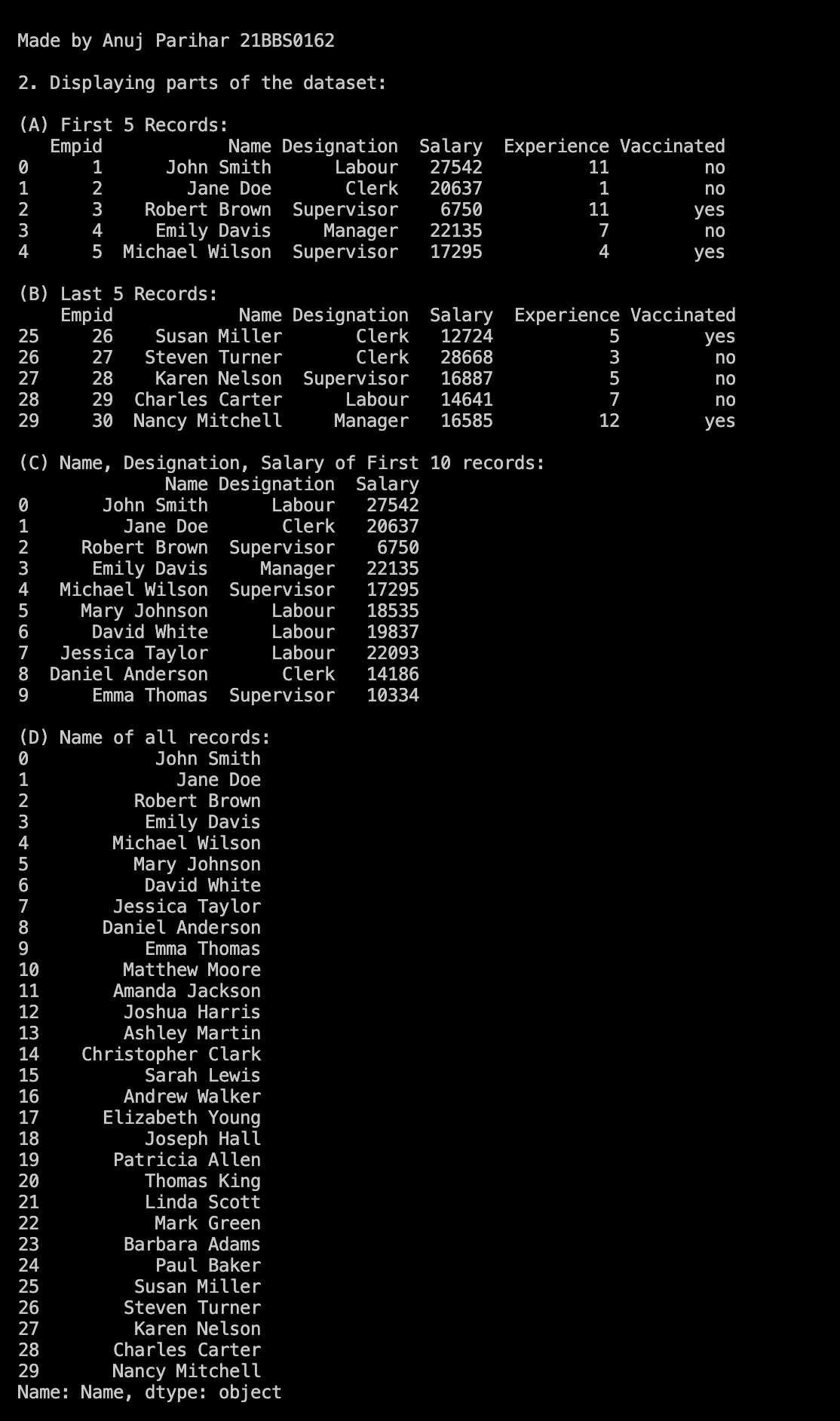
3. Name, Designation, Salary of First 10 records

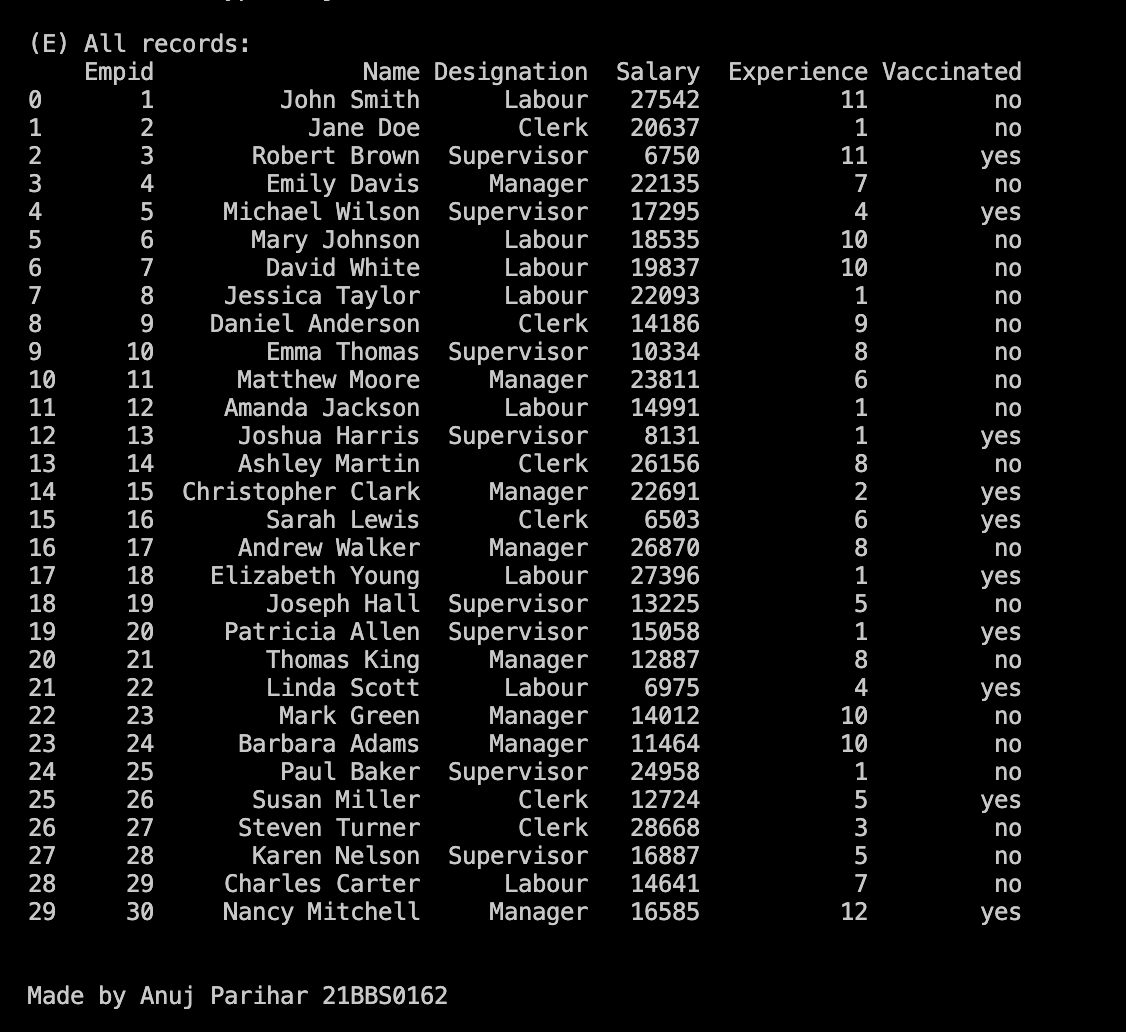
4. Name of all records

5. All records

**Input:** The input is the same as the above for this experiment

**Output:**

****

****

**Results:**

In the above experiment, we get to learn different ways to show different data from a csv

**Aim:** Display the following statistical measures of the dataset

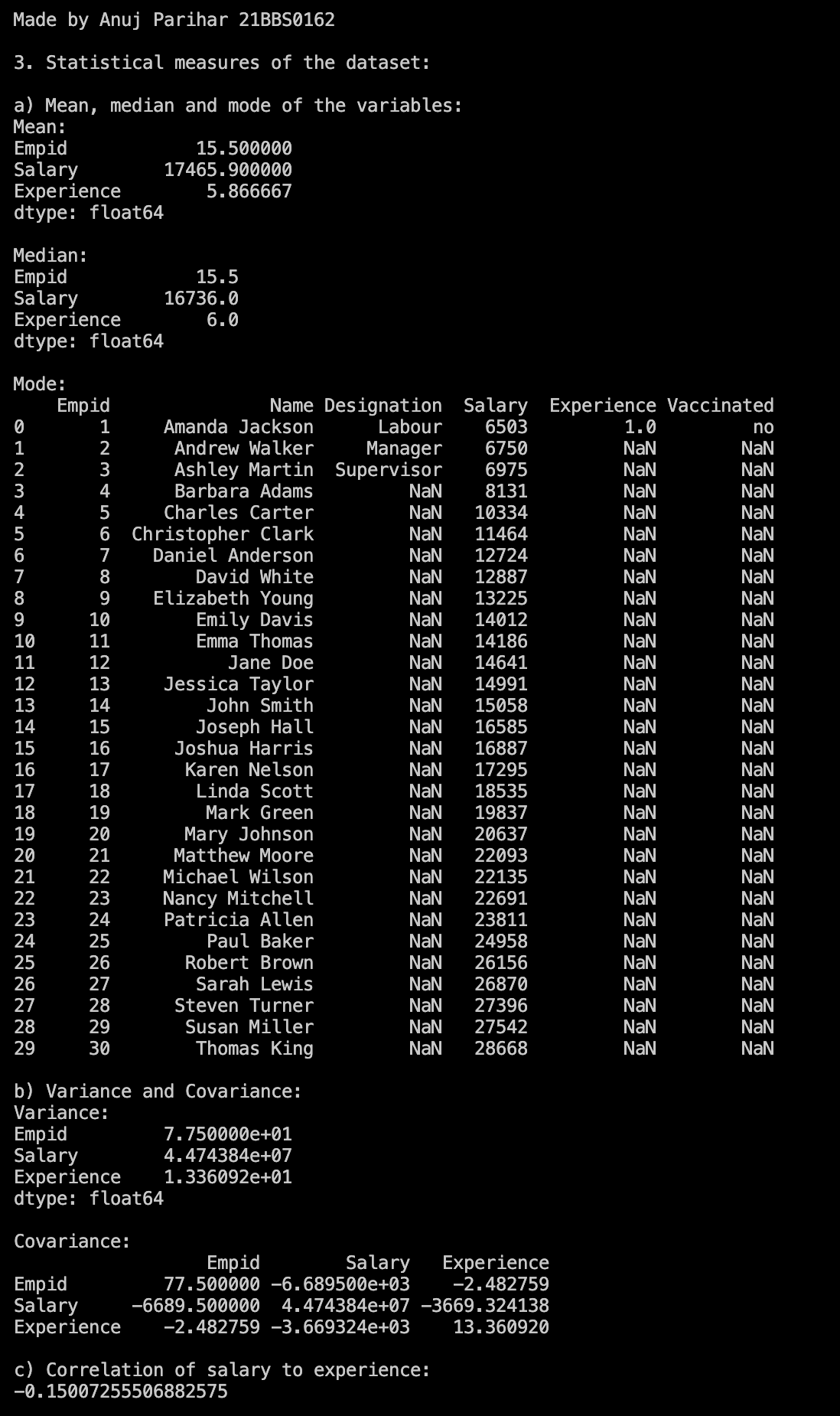
a) Mean, median and mode of the variables

b) Variance and Covariance

c) Correlation of salary to experience

**Input:** The input for the following question is the same as the above

**Output:**

****

**Results:** In the following experiment, we learn how to calculate different statistical measures from a given dataset