

**ANL252 (Online)**

**Python for Data Analytics**

# **Tutor-Marked Assignment**

**July 2022 Presentation**

**Submitted by:**

|  |  |
| --- | --- |
| **Name** | **PI No.** |
| **Tan Wenying Audrey** | **K2170805** |

**Tutorial Group: ­­­­­­­­­­­­ T 09**

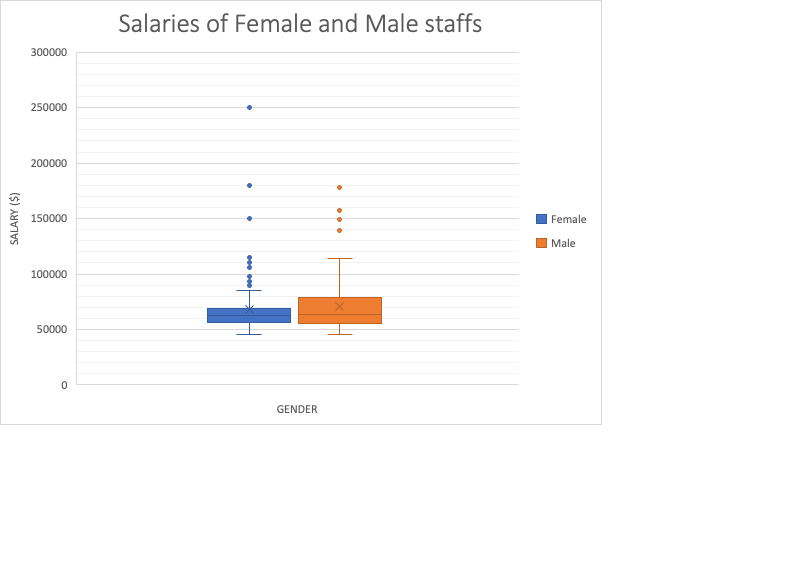
**Instructor’s Name: Dr Munish Kumar**

**Submission Date: 14/08/2022**

**1a)**

**Figure 1**

*Boxplot of salaries of Female and Male staffs*



**Table 1**

*Values extracted from Figure 1*

|  |  |  |
| --- | --- | --- |
| Measure | Female | Male |
| Maximum | 85,028.00 | 113,999.00 |
| Third quartile | 69,254.75 | 78,958.50 |
| Mean | 67,686.53 | 71,063.81 |
| Median | 62,449.50 | 63,799.50 |
| First quartile | 56,010.50 | 55,416.75 |
| Minimum | 45,046.00 | 45,155.00 |

**Table 2**

*Summarised Excel table used to create boxplot in Figure 1*

|  |  |
| --- | --- |
| Female | Male |
| 46335 | 93046 |
| 66825 | 66808 |
| 60446 | 46998 |
| 50825 | 74312 |
| 58062 | 50178 |
| 48413 | 47414 |
| 66593 | 61809 |
| 66738 | 51044 |
| 46799 | 72992 |
| 106367 | 53060 |
| 47837 | 63003 |
| 52984 | 60380 |
| 59026 | 71707 |
| 68099 | 57575 |
| 61555 | 75281 |
| 47961 | 46664 |
| 54670 | 48285 |
| 62061 | 93093 |
| 70131 | 71339 |
| 74241 | 63025 |
| 64995 | 46837 |
| 58275 | 68051 |
| 61844 | 62957 |
| 65729 | 77692 |
| 62514 | 59892 |
| 89292 | 108987 |
| 58370 | 70545 |
| 95920 | 48888 |
| 64971 | 52249 |
| 53018 | 61962 |
| 59238 | 148999 |
| 74226 | 113999 |
| 68999 | 157000 |
| 64816 | 67251 |
| 114800 | 87826 |
| 55425 | 59369 |
| 51777 | 66541 |
| 61242 | 92328 |
| 60270 | 140920 |
| 99351 | 70468 |
| 52057 | 50470 |
| 65714 | 53250 |
| 105688 | 178000 |
| 57859 | 59365 |
| 49256 | 51259 |
| 70621 | 72106 |
| 45069 | 54285 |
| 64066 | 65288 |
| 64955 | 83552 |
| 63025 | 55875 |
| 60724 | 74679 |
| 57815 | 50482 |
| 59124 | 63353 |
| 64786 | 74417 |
| 93554 | 96820 |
| 57583 | 54933 |
| 63430 | 83667 |
| 62385 | 65893 |
| 51505 | 45115 |
| 62910 | 55578 |
| 54005 | 53180 |
| 57748 | 92989 |
| 63763 | 72609 |
| 63322 | 76029 |
| 95660 | 55800 |
| 63813 | 70507 |
| 150290 | 57834 |
| 60627 | 59472 |
| 61242 | 82758 |
| 65902 | 62506 |
| 81584 | 47211 |
| 250000 | 64724 |
| 64991 | 63291 |
| 110000 | 110929 |
| 66149 | 60340 |
| 57568 | 46428 |
| 107226 | 64246 |
| 180000 | 62162 |
| 47434 | 59144 |
| 77915 | 61729 |
| 71860 | 52788 |
| 51920 | 99020 |
| 85028 | 92329 |
| 46430 | 47001 |
| 61422 | 52846 |
| 53171 | 53564 |
| 50923 | 83363 |
| 56147 | 57975 |
| 80512 | 58371 |
| 59728 | 100416 |
| 46738 | 72640 |
| 53492 | 64520 |
| 61349 | 55722 |
| 74326 | 63108 |
| 57815 | 67176 |
| 56339 | 84903 |
| 45433 | 68678 |
| 63676 | 50428 |
| 58530 | 93206 |
| 64919 | 64397 |
| 60120 | 83082 |
| 71966 | 88527 |
| 68182 | 56294 |
| 45395 | 50373 |
| 63000 | 90100 |
| 66074 | 138888 |
| 93396 |  |
| 63878 |  |
| 68407 |  |
| 57954 |  |
| 61154 |  |
| 55965 |  |
| 60070 |  |
| 75188 |  |
| 63381 |  |
| 99280 |  |
| 56847 |  |
| 61656 |  |
| 73330 |  |
| 69340 |  |
| 55688 |  |
| 62068 |  |
| 74669 |  |
| 50750 |  |
| 65707 |  |
| 61584 |  |
| 63478 |  |
| 101199 |  |
| 74813 |  |
| 51908 |  |
| 62810 |  |
| 45998 |  |
| 62659 |  |
| 100031 |  |
| 72202 |  |
| 47750 |  |
| 97999 |  |
| 54828 |  |
| 63682 |  |
| 49773 |  |
| 45046 |  |
| 89883 |  |
| 50274 |  |
| 55315 |  |

**Figure 2**

*Scatterplot of Age and Salary with Units as Hue*

**Table 3**

*Extracting year from JoinDate to determine Age of each staff*

|  |  |  |
| --- | --- | --- |
| Change to date | Year | Age |
| 10/07/1986 | 1986 | 36 |
| 05/25/86 | 1986 | 36 |
| 04/20/85 | 1985 | 37 |
| 06/10/1984 | 1984 | 38 |
| 09/08/1989 | 1989 | 33 |
| 07/30/83 | 1983 | 39 |
| 05/09/1965 | 1965 | 57 |
| 06/11/1970 | 1970 | 52 |
| 03/12/1973 | 1973 | 49 |
| 11/23/85 | 1985 | 37 |
| 01/04/1984 | 1984 | 38 |
| 10/15/84 | 1984 | 38 |
| 04/04/1987 | 1987 | 35 |
| 02/11/1970 | 1970 | 52 |
| 06/03/1967 | 1967 | 55 |
| 10/27/70 | 1970 | 52 |
| 03/10/1970 | 1970 | 52 |
| 01/07/1988 | 1988 | 34 |
| 12/11/1976 | 1976 | 46 |
| 08/27/72 | 1972 | 50 |
| 09/22/89 | 1989 | 33 |
| 08/25/82 | 1982 | 40 |
| 01/12/1974 | 1974 | 48 |
| 10/12/1954 | 1954 | 68 |
| 12/22/70 | 1970 | 52 |
| 07/07/1984 | 1984 | 38 |
| 04/17/66 | 1966 | 56 |
| 08/16/84 | 1984 | 38 |
| 11/24/79 | 1979 | 43 |
| 11/08/1988 | 1988 | 34 |
| 05/07/1992 | 1992 | 30 |
| 09/01/1986 | 1986 | 36 |
| 08/24/83 | 1983 | 39 |
| 02/25/51 | 1951 | 71 |
| 08/29/88 | 1988 | 34 |
| 02/09/1969 | 1969 | 53 |
| 04/19/90 | 1990 | 32 |
| 04/18/80 | 1980 | 42 |
| 05/05/1988 | 1988 | 34 |
| 09/23/73 | 1973 | 49 |
| 02/24/79 | 1979 | 43 |
| 05/07/1965 | 1965 | 57 |
| 05/12/1980 | 1980 | 42 |
| 09/05/1981 | 1981 | 41 |
| 06/18/92 | 1992 | 30 |
| 08/15/68 | 1968 | 54 |
| 03/28/83 | 1983 | 39 |
| 05/21/79 | 1979 | 43 |
| 09/05/1984 | 1984 | 38 |
| 02/24/69 | 1969 | 53 |
| 10/07/1982 | 1982 | 40 |
| 03/10/1979 | 1979 | 43 |
| 08/19/59 | 1959 | 63 |
| 11/11/1989 | 1989 | 33 |
| 12/17/75 | 1975 | 47 |
| 07/11/1981 | 1981 | 41 |
| 05/31/88 | 1988 | 34 |
| 10/23/71 | 1971 | 51 |
| 11/22/66 | 1966 | 56 |
| 09/29/69 | 1969 | 53 |
| 06/10/1986 | 1986 | 36 |
| 02/20/79 | 1979 | 43 |
| 09/14/88 | 1988 | 34 |
| 10/05/1988 | 1988 | 34 |
| 09/11/1972 | 1972 | 50 |
| 07/18/89 | 1989 | 33 |
| 04/16/79 | 1979 | 43 |
| 05/31/74 | 1974 | 48 |
| 09/15/85 | 1985 | 37 |
| 10/22/75 | 1975 | 47 |
| 09/30/75 | 1975 | 47 |
| 11/07/1987 | 1987 | 35 |
| 05/23/91 | 1991 | 31 |
| 10/09/1974 | 1974 | 48 |
| 05/09/1984 | 1984 | 38 |
| 07/18/88 | 1988 | 34 |
| 03/22/66 | 1966 | 56 |
| 08/31/81 | 1981 | 41 |
| 09/19/88 | 1988 | 34 |
| 05/16/83 | 1983 | 39 |
| 11/06/1986 | 1986 | 36 |
| 01/04/1964 | 1964 | 58 |
| 05/08/1980 | 1980 | 42 |
| 08/07/1986 | 1986 | 36 |
| 05/06/1989 | 1989 | 33 |
| 08/27/83 | 1983 | 39 |
| 09/16/84 | 1984 | 38 |
| 11/05/1978 | 1978 | 44 |
| 07/07/1984 | 1984 | 38 |
| 10/05/1986 | 1986 | 36 |
| 08/28/63 | 1963 | 59 |
| 02/08/1970 | 1970 | 52 |
| 11/25/78 | 1978 | 44 |
| 08/25/76 | 1976 | 46 |
| 05/15/70 | 1970 | 52 |
| 09/01/1989 | 1989 | 33 |
| 12/08/1973 | 1973 | 49 |
| 04/14/55 | 1955 | 67 |
| 03/02/1980 | 1980 | 42 |
| 09/08/1969 | 1969 | 53 |
| 05/24/79 | 1979 | 43 |
| 02/10/1976 | 1976 | 46 |
| 07/04/1988 | 1988 | 34 |
| 04/05/1973 | 1973 | 49 |
| 05/21/83 | 1983 | 39 |
| 12/27/88 | 1988 | 34 |
| 05/02/1989 | 1989 | 33 |
| 11/21/72 | 1972 | 50 |
| 12/05/1974 | 1974 | 48 |
| 04/16/81 | 1981 | 41 |
| 04/06/1979 | 1979 | 43 |
| 09/27/87 | 1987 | 35 |
| 07/05/1980 | 1980 | 42 |
| 06/14/87 | 1987 | 35 |
| 09/21/54 | 1954 | 68 |
| 09/27/88 | 1988 | 34 |
| 02/18/83 | 1983 | 39 |
| 11/15/82 | 1982 | 40 |
| 12/03/1976 | 1976 | 46 |
| 08/25/78 | 1978 | 44 |
| 04/04/1986 | 1986 | 36 |
| 12/08/1983 | 1983 | 39 |
| 05/22/77 | 1977 | 45 |
| 07/28/83 | 1983 | 39 |
| 05/02/1978 | 1978 | 44 |
| 03/17/66 | 1966 | 56 |
| 02/14/73 | 1973 | 49 |
| 08/26/86 | 1986 | 36 |
| 07/11/1989 | 1989 | 33 |
| 11/24/89 | 1989 | 33 |
| 01/19/76 | 1976 | 46 |
| 01/28/85 | 1985 | 37 |
| 10/26/81 | 1981 | 41 |
| 12/01/1974 | 1974 | 48 |
| 09/04/1983 | 1983 | 39 |
| 08/09/1974 | 1974 | 48 |
| 05/15/63 | 1963 | 59 |
| 08/10/1981 | 1981 | 41 |
| 05/11/1985 | 1985 | 37 |
| 07/22/82 | 1982 | 40 |
| 05/19/88 | 1988 | 34 |
| 01/18/52 | 1952 | 70 |
| 09/22/70 | 1970 | 52 |
| 01/07/1985 | 1985 | 37 |
| 07/03/1972 | 1972 | 50 |
| 12/02/1983 | 1983 | 39 |
| 03/10/1975 | 1975 | 47 |
| 03/18/87 | 1987 | 35 |
| 05/06/1983 | 1983 | 39 |
| 10/01/1981 | 1981 | 41 |
| 03/31/69 | 1969 | 53 |
| 06/29/85 | 1985 | 37 |
| 07/10/1988 | 1988 | 34 |
| 11/14/55 | 1955 | 67 |
| 11/07/1958 | 1958 | 64 |
| 10/18/81 | 1981 | 41 |
| 06/19/61 | 1961 | 61 |
| 07/01/1972 | 1972 | 50 |
| 07/10/1983 | 1983 | 39 |
| 02/21/74 | 1974 | 48 |
| 03/06/1988 | 1988 | 34 |
| 07/01/1984 | 1984 | 38 |
| 10/02/1969 | 1969 | 53 |
| 01/12/1973 | 1973 | 49 |
| 08/24/90 | 1990 | 32 |
| 11/07/1974 | 1974 | 48 |
| 06/01/1964 | 1964 | 58 |
| 02/09/1972 | 1972 | 50 |
| 09/02/1983 | 1983 | 39 |
| 01/16/67 | 1967 | 55 |
| 06/05/1967 | 1967 | 55 |
| 10/08/1970 | 1970 | 52 |
| 01/17/79 | 1979 | 43 |
| 03/16/81 | 1981 | 41 |
| 12/27/58 | 1958 | 64 |
| 05/12/1989 | 1989 | 33 |
| 02/11/1952 | 1952 | 70 |
| 12/21/74 | 1974 | 48 |
| 08/10/1988 | 1988 | 34 |
| 08/19/77 | 1977 | 45 |
| 08/12/1979 | 1979 | 43 |
| 09/22/76 | 1976 | 46 |
| 12/31/84 | 1984 | 38 |
| 11/09/1972 | 1972 | 50 |
| 07/07/1986 | 1986 | 36 |
| 06/30/89 | 1989 | 33 |
| 11/24/87 | 1987 | 35 |
| 07/25/79 | 1979 | 43 |
| 09/09/1965 | 1965 | 57 |
| 11/23/81 | 1981 | 41 |
| 04/05/1987 | 1987 | 35 |
| 11/25/87 | 1987 | 35 |
| 12/10/1986 | 1986 | 36 |
| 02/02/1983 | 1983 | 39 |
| 08/17/86 | 1986 | 36 |
| 03/17/88 | 1988 | 34 |
| 04/17/86 | 1986 | 36 |
| 11/06/1984 | 1984 | 38 |
| 11/08/1983 | 1983 | 39 |
| 04/26/86 | 1986 | 36 |
| 09/05/1991 | 1991 | 31 |
| 11/28/73 | 1973 | 49 |
| 03/31/77 | 1977 | 45 |
| 05/15/87 | 1987 | 35 |
| 08/25/89 | 1989 | 33 |
| 05/24/87 | 1987 | 35 |
| 01/02/1951 | 1951 | 71 |
| 03/28/82 | 1982 | 40 |
| 03/11/1984 | 1984 | 38 |
| 09/22/76 | 1976 | 46 |
| 04/06/1985 | 1985 | 37 |
| 05/09/1977 | 1977 | 45 |
| 04/14/81 | 1981 | 41 |
| 07/06/1968 | 1968 | 54 |
| 06/14/83 | 1983 | 39 |
| 08/09/1983 | 1983 | 39 |
| 04/26/84 | 1984 | 38 |
| 12/02/1978 | 1978 | 44 |
| 10/31/77 | 1977 | 45 |
| 09/08/1977 | 1977 | 45 |
| 02/16/84 | 1984 | 38 |
| 07/05/1979 | 1979 | 43 |
| 03/15/85 | 1985 | 37 |
| 09/16/75 | 1975 | 47 |
| 07/08/1981 | 1981 | 41 |
| 01/28/91 | 1991 | 31 |
| 01/07/1987 | 1987 | 35 |
| 07/20/86 | 1986 | 36 |
| 09/05/1985 | 1985 | 37 |
| 05/30/68 | 1968 | 54 |
| 06/06/1986 | 1986 | 36 |
| 01/07/1974 | 1974 | 48 |
| 05/24/53 | 1953 | 69 |
| 06/06/1968 | 1968 | 54 |
| 02/21/84 | 1984 | 38 |
| 03/23/77 | 1977 | 45 |
| 04/23/86 | 1986 | 36 |
| 03/28/78 | 1978 | 44 |
| 01/15/68 | 1968 | 54 |
| 11/15/76 | 1976 | 46 |
| 06/03/1986 | 1986 | 36 |
| 12/17/87 | 1987 | 35 |
| 09/14/79 | 1979 | 43 |
| 08/17/78 | 1978 | 44 |
| 08/26/80 | 1980 | 42 |
| 10/11/1981 | 1981 | 41 |
| 10/24/87 | 1987 | 35 |
| 07/09/1970 | 1970 | 52 |
| 08/02/1980 | 1980 | 42 |
| 05/21/87 | 1987 | 35 |

**Table 4**

*Summarised Excel table used to create scatterplot in Figure 2*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Salary | Unit | Admin | C-Level | Engineering | IT | Manufacturing | Sales |
| 36 | 46335 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46335 | #N/A |
| 36 | 66825 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 66825 | #N/A |
| 37 | 60446 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60446 | #N/A |
| 38 | 93046 | Admin | 93046 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 33 | 50825 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50825 | #N/A |
| 39 | 58062 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 58062 | #N/A |
| 57 | 48413 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 48413 | #N/A |
| 52 | 66808 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 66808 |
| 49 | 66593 | IT | #N/A | #N/A | #N/A | 66593 | #N/A | #N/A |
| 37 | 66738 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 66738 | #N/A |
| 38 | 46998 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46998 | #N/A |
| 38 | 46799 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46799 | #N/A |
| 35 | 106367 | Admin | 106367 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 52 | 47837 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47837 | #N/A |
| 55 | 52984 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 52984 | #N/A |
| 52 | 59026 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59026 | #N/A |
| 52 | 74312 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 74312 | #N/A |
| 34 | 50178 | IT | #N/A | #N/A | #N/A | 50178 | #N/A | #N/A |
| 46 | 47414 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47414 | #N/A |
| 50 | 68099 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 68099 | #N/A |
| 33 | 61555 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 61555 |
| 40 | 47961 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47961 | #N/A |
| 48 | 54670 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 54670 | #N/A |
| 68 | 61809 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 61809 |
| 52 | 51044 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 51044 | #N/A |
| 38 | 62061 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62061 | #N/A |
| 56 | 70131 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 70131 | #N/A |
| 38 | 72992 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 72992 |
| 43 | 53060 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53060 | #N/A |
| 34 | 74241 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 74241 |
| 30 | 64995 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64995 | #N/A |
| 36 | 63003 | Admin | 63003 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 39 | 60380 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60380 | #N/A |
| 71 | 58275 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 58275 | #N/A |
| 34 | 61844 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 61844 |
| 53 | 71707 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 71707 |
| 32 | 65729 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 65729 |
| 42 | 57575 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57575 | #N/A |
| 34 | 75281 | IT | #N/A | #N/A | #N/A | 75281 | #N/A | #N/A |
| 49 | 62514 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62514 | #N/A |
| 43 | 89292 | IT | #N/A | #N/A | #N/A | 89292 | #N/A | #N/A |
| 57 | 58370 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 58370 |
| 42 | 95920 | IT | #N/A | #N/A | #N/A | 95920 | #N/A | #N/A |
| 41 | 64971 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64971 | #N/A |
| 30 | 53018 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53018 | #N/A |
| 54 | 59238 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59238 | #N/A |
| 39 | 46664 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46664 | #N/A |
| 43 | 48285 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 48285 | #N/A |
| 38 | 93093 | IT | #N/A | #N/A | #N/A | 93093 | #N/A | #N/A |
| 53 | 71339 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 71339 |
| 40 | 63025 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63025 | #N/A |
| 43 | 74226 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 74226 | #N/A |
| 63 | 46837 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46837 | #N/A |
| 33 | 68999 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 68999 |
| 47 | 68051 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 68051 | #N/A |
| 41 | 62957 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62957 | #N/A |
| 34 | 64816 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64816 | #N/A |
| 51 | 114800 | IT | #N/A | #N/A | #N/A | 114800 | #N/A | #N/A |
| 56 | 77692 | Engineering | #N/A | #N/A | 77692 | #N/A | #N/A | #N/A |
| 53 | 59892 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59892 | #N/A |
| 36 | 55425 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55425 | #N/A |
| 43 | 108987 | Engineering | #N/A | #N/A | 108987 | #N/A | #N/A | #N/A |
| 34 | 70545 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 70545 |
| 34 | 51777 | IT | #N/A | #N/A | #N/A | 51777 | #N/A | #N/A |
| 50 | 61242 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61242 | #N/A |
| 33 | 60270 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60270 | #N/A |
| 43 | 99351 | Admin | 99351 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 48 | 48888 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 48888 | #N/A |
| 37 | 52249 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 52249 | #N/A |
| 47 | 52057 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 52057 | #N/A |
| 47 | 65714 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 65714 | #N/A |
| 35 | 105688 | Engineering | #N/A | #N/A | 105688 | #N/A | #N/A | #N/A |
| 31 | 57859 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 57859 |
| 48 | 49256 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 49256 | #N/A |
| 38 | 61962 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61962 | #N/A |
| 34 | 70621 | IT | #N/A | #N/A | #N/A | 70621 | #N/A | #N/A |
| 56 | 45069 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45069 | #N/A |
| 41 | 64066 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64066 | #N/A |
| 34 | 64955 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64955 | #N/A |
| 39 | 63025 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63025 | #N/A |
| 36 | 60724 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60724 | #N/A |
| 58 | 148999 | IT | #N/A | #N/A | #N/A | 148999 | #N/A | #N/A |
| 42 | 57815 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57815 | #N/A |
| 36 | 113999 | IT | #N/A | #N/A | #N/A | 113999 | #N/A | #N/A |
| 33 | 59124 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59124 | #N/A |
| 39 | 64786 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64786 | #N/A |
| 38 | 93554 | IT | #N/A | #N/A | #N/A | 93554 | #N/A | #N/A |
| 44 | 57583 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57583 | #N/A |
| 38 | 63430 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63430 | #N/A |
| 36 | 157000 | IT | #N/A | #N/A | #N/A | 157000 | #N/A | #N/A |
| 59 | 67251 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 67251 |
| 52 | 87826 | IT | #N/A | #N/A | #N/A | 87826 | #N/A | #N/A |
| 44 | 59369 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59369 | #N/A |
| 46 | 62385 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62385 | #N/A |
| 52 | 51505 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 51505 | #N/A |
| 33 | 62910 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62910 | #N/A |
| 49 | 54005 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 54005 | #N/A |
| 67 | 57748 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57748 | #N/A |
| 42 | 63763 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63763 | #N/A |
| 53 | 63322 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63322 | #N/A |
| 43 | 95660 | Engineering | #N/A | #N/A | 95660 | #N/A | #N/A | #N/A |
| 46 | 66541 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 66541 | #N/A |
| 34 | 92328 | IT | #N/A | #N/A | #N/A | 92328 | #N/A | #N/A |
| 49 | 140920 | IT | #N/A | #N/A | #N/A | 140920 | #N/A | #N/A |
| 39 | 63813 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63813 | #N/A |
| 34 | 70468 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 70468 |
| 33 | 50470 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50470 | #N/A |
| 50 | 150290 | IT | #N/A | #N/A | #N/A | 150290 | #N/A | #N/A |
| 48 | 60627 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60627 | #N/A |
| 41 | 61242 | IT | #N/A | #N/A | #N/A | 61242 | #N/A | #N/A |
| 43 | 53250 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53250 | #N/A |
| 35 | 65902 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 65902 | #N/A |
| 42 | 178000 | IT | #N/A | #N/A | #N/A | 178000 | #N/A | #N/A |
| 35 | 81584 | IT | #N/A | #N/A | #N/A | 81584 | #N/A | #N/A |
| 68 | 250000 | C-Level | #N/A | 250000 | #N/A | #N/A | #N/A | #N/A |
| 34 | 64991 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64991 | #N/A |
| 39 | 59365 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59365 | #N/A |
| 40 | 51259 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 51259 | #N/A |
| 46 | 72106 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 72106 | #N/A |
| 44 | 54285 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 54285 | #N/A |
| 36 | 110000 | IT | #N/A | #N/A | #N/A | 110000 | #N/A | #N/A |
| 39 | 66149 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 66149 | #N/A |
| 45 | 57568 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57568 | #N/A |
| 39 | 65288 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 65288 | #N/A |
| 44 | 107226 | IT | #N/A | #N/A | #N/A | 107226 | #N/A | #N/A |
| 56 | 180000 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 180000 |
| 49 | 47434 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47434 | #N/A |
| 36 | 83552 | IT | #N/A | #N/A | #N/A | 83552 | #N/A | #N/A |
| 33 | 55875 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 55875 |
| 33 | 74679 | IT | #N/A | #N/A | #N/A | 74679 | #N/A | #N/A |
| 46 | 50482 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50482 | #N/A |
| 37 | 63353 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63353 | #N/A |
| 41 | 77915 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 77915 | #N/A |
| 48 | 74417 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 74417 | #N/A |
| 39 | 96820 | IT | #N/A | #N/A | #N/A | 96820 | #N/A | #N/A |
| 48 | 54933 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 54933 | #N/A |
| 59 | 71860 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 71860 |
| 41 | 83667 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 83667 | #N/A |
| 37 | 65893 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 65893 | #N/A |
| 40 | 45115 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45115 | #N/A |
| 34 | 51920 | Admin | 51920 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 70 | 85028 | IT | #N/A | #N/A | #N/A | 85028 | #N/A | #N/A |
| 52 | 46430 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46430 | #N/A |
| 37 | 61422 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61422 | #N/A |
| 50 | 55578 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55578 | #N/A |
| 39 | 53171 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53171 | #N/A |
| 47 | 50923 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50923 | #N/A |
| 35 | 53180 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53180 | #N/A |
| 39 | 92989 | Engineering | #N/A | #N/A | 92989 | #N/A | #N/A | #N/A |
| 41 | 72609 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 72609 | #N/A |
| 53 | 76029 | IT | #N/A | #N/A | #N/A | 76029 | #N/A | #N/A |
| 37 | 55800 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55800 | #N/A |
| 34 | 56147 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 56147 | #N/A |
| 67 | 80512 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 80512 | #N/A |
| 64 | 70507 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 70507 | #N/A |
| 41 | 57834 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57834 | #N/A |
| 61 | 59472 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59472 | #N/A |
| 50 | 82758 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 82758 | #N/A |
| 39 | 62506 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62506 | #N/A |
| 48 | 47211 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47211 | #N/A |
| 34 | 64724 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64724 | #N/A |
| 38 | 63291 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 63291 |
| 53 | 59728 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59728 | #N/A |
| 49 | 46738 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46738 | #N/A |
| 32 | 53492 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53492 | #N/A |
| 48 | 61349 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61349 | #N/A |
| 58 | 74326 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 74326 |
| 50 | 110929 | IT | #N/A | #N/A | #N/A | 110929 | #N/A | #N/A |
| 39 | 60340 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60340 | #N/A |
| 55 | 57815 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57815 | #N/A |
| 55 | 56339 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 56339 | #N/A |
| 52 | 45433 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45433 | #N/A |
| 43 | 63676 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63676 | #N/A |
| 41 | 58530 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 58530 | #N/A |
| 64 | 64919 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64919 | #N/A |
| 33 | 60120 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 60120 |
| 70 | 71966 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 71966 | #N/A |
| 48 | 46428 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 46428 | #N/A |
| 34 | 64246 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 64246 | #N/A |
| 45 | 62162 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62162 | #N/A |
| 43 | 59144 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 59144 | #N/A |
| 46 | 68182 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 68182 | #N/A |
| 38 | 61729 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61729 | #N/A |
| 50 | 52788 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 52788 | #N/A |
| 36 | 45395 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45395 | #N/A |
| 33 | 99020 | IT | #N/A | #N/A | #N/A | 99020 | #N/A | #N/A |
| 35 | 63000 | Admin | 63000 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 43 | 66074 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 66074 | #N/A |
| 57 | 92329 | IT | #N/A | #N/A | #N/A | 92329 | #N/A | #N/A |
| 41 | 47001 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47001 | #N/A |
| 35 | 93396 | Engineering | #N/A | #N/A | 93396 | #N/A | #N/A | #N/A |
| 35 | 63878 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63878 | #N/A |
| 36 | 68407 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 68407 | #N/A |
| 39 | 52846 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 52846 | #N/A |
| 36 | 57954 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57954 | #N/A |
| 34 | 53564 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 53564 | #N/A |
| 36 | 61154 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61154 | #N/A |
| 38 | 83363 | Engineering | #N/A | #N/A | 83363 | #N/A | #N/A | #N/A |
| 39 | 55965 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55965 | #N/A |
| 36 | 57975 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 57975 | #N/A |
| 31 | 60070 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 60070 | #N/A |
| 49 | 75188 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 75188 | #N/A |
| 45 | 63381 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63381 | #N/A |
| 35 | 99280 | Engineering | #N/A | #N/A | 99280 | #N/A | #N/A | #N/A |
| 33 | 56847 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 56847 | #N/A |
| 35 | 58371 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 58371 | #N/A |
| 71 | 61656 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61656 | #N/A |
| 40 | 73330 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 73330 | #N/A |
| 38 | 69340 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 69340 | #N/A |
| 46 | 55688 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55688 | #N/A |
| 37 | 62068 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62068 | #N/A |
| 45 | 74669 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 74669 | #N/A |
| 41 | 50750 | IT | #N/A | #N/A | #N/A | 50750 | #N/A | #N/A |
| 54 | 65707 | IT | #N/A | #N/A | #N/A | 65707 | #N/A | #N/A |
| 39 | 100416 | Engineering | #N/A | #N/A | 100416 | #N/A | #N/A | #N/A |
| 39 | 72640 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 72640 | #N/A |
| 38 | 64520 | Admin | 64520 | #N/A | #N/A | #N/A | #N/A | #N/A |
| 44 | 61584 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 61584 | #N/A |
| 45 | 55722 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55722 | #N/A |
| 45 | 63108 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63108 | #N/A |
| 38 | 63478 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63478 | #N/A |
| 43 | 101199 | Engineering | #N/A | #N/A | 101199 | #N/A | #N/A | #N/A |
| 37 | 74813 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 74813 | #N/A |
| 47 | 67176 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 67176 | #N/A |
| 41 | 84903 | IT | #N/A | #N/A | #N/A | 84903 | #N/A | #N/A |
| 31 | 51908 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 51908 | #N/A |
| 35 | 62810 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62810 | #N/A |
| 36 | 45998 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45998 | #N/A |
| 37 | 68678 | IT | #N/A | #N/A | #N/A | 68678 | #N/A | #N/A |
| 54 | 62659 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 62659 | #N/A |
| 36 | 100031 | IT | #N/A | #N/A | #N/A | 100031 | #N/A | #N/A |
| 48 | 50428 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50428 | #N/A |
| 69 | 72202 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 72202 | #N/A |
| 54 | 47750 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 47750 | #N/A |
| 38 | 97999 | IT | #N/A | #N/A | #N/A | 97999 | #N/A | #N/A |
| 45 | 54828 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 54828 | #N/A |
| 36 | 93206 | IT | #N/A | #N/A | #N/A | 93206 | #N/A | #N/A |
| 44 | 63682 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 63682 | #N/A |
| 54 | 64397 | Sales | #N/A | #N/A | #N/A | #N/A | #N/A | 64397 |
| 46 | 83082 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 83082 | #N/A |
| 36 | 49773 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 49773 | #N/A |
| 35 | 88527 | IT | #N/A | #N/A | #N/A | 88527 | #N/A | #N/A |
| 43 | 56294 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 56294 | #N/A |
| 44 | 45046 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 45046 | #N/A |
| 42 | 50373 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50373 | #N/A |
| 41 | 89883 | IT | #N/A | #N/A | #N/A | 89883 | #N/A | #N/A |
| 35 | 90100 | IT | #N/A | #N/A | #N/A | 90100 | #N/A | #N/A |
| 52 | 138888 | IT | #N/A | #N/A | #N/A | 138888 | #N/A | #N/A |
| 42 | 50274 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 50274 | #N/A |
| 35 | 55315 | Manufacturing | #N/A | #N/A | #N/A | #N/A | 55315 | #N/A |

**Explanation of how Table 2 was derived in Excel**

Table 2 was derived byfiltering column ‘Gender’ by ‘F’ and retrieving the data as a column then doing likewise for ‘M’. Then all the columns were selected to insert a boxplot shown in Figure 1.

**Explanation of how Table 3 was derived in Excel**

Table 3 column named ‘Change to date’ was derived by using a formula “=TEXT(C2,"dd/mm/yyyy")” to convert the dates in column ‘BirthYear’ to the date then month then year format. Column ‘Year’ was derived by using a formula “=CONCATENATE(19,RIGHT(P2,2))” to concatenate number 19 to each of the last two digit per cell in ‘Change to date’ to make up the year. Column named ‘Age’ was derived by using a formula “=2022-Q2”. This means the year 2022 was subtracted from each of the cells in the ‘Year’ column.

**Explanation of how Table 4 was derived in Excel**

Table 4 was derived by extracting the ‘Age’ in Table 3 and extracting ‘Salary’ and ‘Unit’ from the original TMA\_Data. The new columns called ‘Admin’, ‘C-Level’, ‘Engineering’, ‘IT’, ‘Manufacturing’ and ‘Sales’ were derived using Excel formulas of ‘If’. For example, if the text in the ‘Unit’ column’s call shows the same text as the column title named ‘Admin’, the cell in the ‘Admin’ column should show the salary, otherwise it will show ‘NA’. The formula used for the ‘Admin’ column was “=IF(C2=$D$1,B2,NA())”. This logic was then applied to the other five columns. Columns ‘Age’, ‘Admin’, ‘C-Level’, ‘Engineering’, ‘IT’, ‘Manufacturing’ and ‘Sales’ were then selected to insert the Figure 2 scatterplot in Excel.

**Insights from analysis**

Figure 1 shows that females earn less than males. This is seen by the data in Table 1 extracted from the boxplot. Comparison of the two boxplot’s top and bottom whiskers which represents the maximum and minimum salary, the “X” in the boxplot which represents the mean, and the horizontal line which represents the median, can be made. From the comparison, females have lower figures for each comparison as compared to males. In addition, the box for females is shorter than that for males, and the whiskers are shorter for females than for males. These show that the salary varies less for females than males. Another observation is that there are more outliers for females than males. Also, the staff earning the highest salary at $250,000 is a female that is an outlier.

To glean more insight on the salary, a scatterplot in Figure 2 can be plotted. The highest income earner shown in orange datapoint is working in the unit called C-Level and aged between 60 to 70 years old. This earner is also a female earning around $250,000 from the data seen in Figure 1. The lowest income earners are in blue datapoints, they earn less than $50,000, come from the Manufacturing unit, and are aged between their 30s to 60s. In general, C-Level personnel earn the most followed by IT personnel, Engineers, Sales then Manufacturing personnel.

The staff composition can also be observed from the scatterplot. The multiple blue datapoints in the scatterplot shows that most of the staffs are aged between 30 to 50 years old and are working in the manufacturing unit. There are very few C-Level, Admin and Engineering staffs as seen from the few orange, purple and pink datapoints.

From the scatterplot and boxplot, there was one sales personnel who is an outlier by earning between $150,000 to $200,000. The majority of other sales personnel take in earnings of between $50,000 to $100,000, which is much lower. This could be a data error due to the vast differences in the earnings. Alternatively, the sales personnel may have performed well to earn more or that the company values this person by giving a larger pay out. The python user can then further check with the Human Resource department on the accuracy of the data after finding out these information. Removal or editing of those data that might have been incorrectly keyed in might be required if errors are found.

**1b)**

Figure 5 and Figure 6 shows the first summarised table and first chart.

Figure 9 and Figure 10 shows the second summarised table and second chart.

**Python code (in text):**

**Libraries**

#import pandas

import pandas as pd

#import matplotlib

import matplotlib.pyplot as plt

import matplotlib.patches as mpatches

#import seaborn

import seaborn as sns

#import ast

import ast

#import datetime

from datetime import date, time, datetime

#import numpy

import numpy as np

**Import Excel**

#Read excel

df = pd.read\_csv('TMA\_Data.csv')

### **Display first 5 rows of dataframe and check type of data**

#Display the data in the first 5 rows of dataframe

df.head()

#Print type function to check type

print(df.dtypes)

**Preparing data for chart 1 on boxplot**

#Extract all the rows with Gender = F or female

contain\_values = df[df['Gender'].str.contains('F')]

#For these rows with Gender =F, filter out the Salary

contain\_values = contain\_values.filter(['Salary'])

#Rename the new column

contain\_values=contain\_values.rename(columns={'Salary':'Salary\_of\_Females'})

#Extract all the rows with Gender = M or Male

contain\_values2 = df[df['Gender'].str.contains('M')]

#For these rows with Gender =M, filter out the Salary

contain\_values2 = contain\_values2.filter(['Salary'])

#Rename the new column

contain\_values2=contain\_values2.rename(columns={'Salary':'Salary\_of\_Males'})

**Summarised table**

#create boxplot summarised table

boxplot\_table = pd.concat([contain\_values, contain\_values2], axis=1)

#Print out the first and last five observations in the dataframe

display(boxplot\_table)

**Chart 1-Boxplot**

#Combining data of salaries from M and F into numpy array

data = np.array([contain\_values, contain\_values2], dtype=object)

#Set the size of the plot

fig = plt.figure(figsize =(5, 3))

#Creating axes instance

ax = fig.add\_axes([0, 0, 1, 1])

#Creating plot

bp = ax.boxplot(data)

#X and Y Labels

plt.xlabel('Gender')

plt.ylabel('Salary ($)')

#Adding title

plt.title("Salaries of Female and Male staffs")

#Adding ticks

plt.xticks([1, 2], ['Female', 'Male'])

#Plot the boxplot

box = plt.boxplot(data, #array to be plotted

patch\_artist=True, #fill with color

flierprops={'markeredgecolor': 'None'}, #no marker edger for outliers

showmeans=True, #show the mean

meanprops={"marker":"x","markerfacecolor":"black", "markeredgecolor":"black"}) #set the marker type and colour of the mean

#Fill the outliers with two colours

cols = ['steelblue', 'chocolate']

for f, fc in zip(box['fliers'], cols):

f.set\_markerfacecolor(fc)

#Fill the boxplot with two colours

colors = ['steelblue', 'chocolate']

for patch, color in zip(box['boxes'], colors):

patch.set\_facecolor(color)

#set legend

orange\_patch=mpatches.Patch(color='steelblue',label="Female") #Set colour for Females in legend

blue\_patch=mpatches.Patch(color='chocolate',label="Male") #Set colour for Males in legend

plt.legend(loc='center right', bbox\_to\_anchor=(1.25, 0.5), ncol=1, handles=[orange\_patch, blue\_patch])

#show plot

plt.show()

**Preparing data for chart 2 on scatterplot**

#Extract last two digits of BirthYear

df['Year'] = df['BirthYear'].str[-2:]

#Create a new column starting with 19 to show the year starting from 1900s

df['frontyear'] = 19

#Create a new column of dates

df["fullyear"] = df["frontyear"].astype(str)+ df["Year"].astype(str)

#Create a new column for this year of 2022

df["thisyear"] = 2022

#Get the age by taking this year substracted by the DOB

df["Age"] = df["thisyear"]- df["fullyear"].astype(int)

df.head()

**Summarised table**

#Create scatterplot summarised table

scatterplot\_table= df[["Age","Salary","Unit"]]

#Display the first and last five observations of the dataframe

display(scatterplot\_table)

#Insert column headers named 'Admin ','C-Level', 'Engineering', 'IT', 'Manufacturing', 'Sales' into the respective columns indicated by the index

#Use loc attribute to return the salary when the data in each row of the Unit column matches each header

scatterplot\_table.insert(3, 'Admin ', scatterplot\_table.loc[scatterplot\_table.Unit=='Admin ','Salary'])

scatterplot\_table.insert(4, 'C-Level', scatterplot\_table.loc[scatterplot\_table.Unit=='C-Level','Salary'])

scatterplot\_table.insert(5, 'Engineering', scatterplot\_table.loc[scatterplot\_table.Unit=='Engineering','Salary'])

scatterplot\_table.insert(6,'IT', scatterplot\_table.loc[scatterplot\_table.Unit=='IT','Salary'])

scatterplot\_table.insert(7,'Manufacturing', scatterplot\_table.loc[scatterplot\_table.Unit=='Manufacturing','Salary'])

scatterplot\_table.insert(8,'Sales', scatterplot\_table.loc[scatterplot\_table.Unit=='Sales','Salary'])

display(scatterplot\_table)

**Chart 2-Scatterplot**

#Plot scatterplot with Seaborn

#To set the colours per Unit

colour\_dict = dict({'Admin ':'blueviolet',

'C-Level':'orange',

'Engineering': 'pink',

'IT': 'yellow',

'Manufacturing': 'lightblue',

'Sales': 'green'})

#Set the order in the legend

orders=['Admin ','C-Level','Engineering', 'IT', 'Manufacturing', 'Sales']

#To set the side of the scatterplot

sns.set(rc={"figure.figsize":(8, 4)})

#To display the scatterplot

g =sns.scatterplot(x="Age", y="Salary",

hue="Unit",

data=df,

palette=colour\_dict,

hue\_order=orders)

#Rename axis

plt.xlabel('Age (In years)')

plt.ylabel('Salary ($)')

#To start the x and y axis at point 0

plt.xlim(0)

plt.ylim(0)

#To set the title

plt.title('Scatterplot of Age against Salary with Units as Hue')

#To set the legend location of scatterplot

g.legend(loc='center left', bbox\_to\_anchor=(1, 0.5), ncol=1)

**Python output:**

**Figure 3**

*Display data in the first 5 rows of the dataframe*

**Graphical user interface, text

Description automatically generated with medium confidence**

**Figure 4**

*Print type function to check type*

**Table

Description automatically generated with low confidence**

**Figure 5**

*Summarised table to view the salary of the categorised Females and Males for each of the 250 staffs.*

**Table

Description automatically generated**

**Figure 6**

*Boxplot plotted using Matplotlib*

**Chart, box and whisker chart

Description automatically generated**

**Figure 7**

*The right most view of the first 5 rows of the dataframe*

**Table

Description automatically generated**

**Figure 8**

*Summarised table of data in ‘Age’, ‘Salary’ and ‘Unit’ used to plot scatterplot*

**Table

Description automatically generated**

**Figure 9**

*Summarised table of data to match Table 4*

**Table

Description automatically generated**

**Figure 10**

*Scatterplot plotted using Seaborn and legend plotted using Matplotlib*

**Chart, scatter chart

Description automatically generated**

**Explanation:**

Please see the comments for the codes.

**1c)**

**Python code (in text):**

#Counting the number of missing values in each column

df.isnull().sum()

#Create a new column with the filled empty rows in the column ‘LeftDate’ with 1 May 2022.

df['LeftDatenew'] = df['LeftDate'].fillna(value=pd.to\_datetime('05-01-2022'))

df.head(5)

#Convert the two columns to datetime format

df['JoinDatenew'] = pd.to\_datetime(df['JoinDate'])

df['LeftDatenew'] = pd.to\_datetime(df['LeftDatenew'])

#Calculate the length of service of each staff

df['length\_of\_service'] = df['LeftDatenew'].dt.year - df['JoinDatenew'].dt.year

df.head(5)

#Calculate the mean, minimum and maximum value from the length of service and round the data to one decimal place

result = round(df.agg({'length\_of\_service': ['mean', 'min', 'max']}),1)

print("Mean, min, and max values of the 250 staffs' length of service:")

print(result)

**Python output:**

**Figure 11**

*Output by counting the number of missing values in each column*

**Table

Description automatically generated**

**Figure 12**

*Right most view of first 5 rows of dataframe. Fill the empty rows in the column ‘LeftDate’ with 1 May 2022 in new column named ‘LeftDatenew’.*

Background pattern

Description automatically generated

**Figure 13**

*Convert columns ‘LeftDatenew’ and ‘JoinDatenew’ to datetime format and calculate the length of service of each staff in a new column called ‘length\_of\_service’*

Background pattern

Description automatically generated

**Figure 14**

*Output of the minimum, maximum and average length of service, expressing in years, rounded to 1 decimal place*

A picture containing text

Description automatically generated

**Explanation:**

Display the number of missing values in every column by using the isnull() function. Fill the empty rows in the column ‘LeftDate’ with 1 May 2022 that has been converted to datetime format. Earlier in question 1b, df.dtypes attribute was used to find the data type of each column in the dataframe ‘JoinDate’ and ‘LeftDate’. The data type for these two columns were ‘object’ and not ‘datetime64[ns]’. Convert the data in the columns ‘JoinDatenew’ and ‘LeftDatenew’ to datetime object. Calculate the length of service of each staff by using the ‘-‘ operator to subtract data in column ‘LeftDatenew’ with data in column ‘JoinDatenew’. These columns also uses dt.year attribute to return the year of the datetime. Use the agg() function to pass a list of functions of mean, min and max to the column ‘length\_of\_service’. This agg() function is enclosed by the round() function to round the figures off to 1 decimal place and saved in ‘result’ variable. The descriptive text and result is then printed.

**1d)**

**Python codes (in text):**

#Create a loop for user input on the full name of the staff

while True:

xyz = input('\nPlease enter the full name of person or type "exit" to exit this query: ')

Staff\_count=df['Staff'].str.fullmatch(xyz).sum() # When the staff's full name keyed in by user is a full match, then the output is integer of more than 0, else it is 0.

if xyz == 'exit': #If user keys in "exit" the query will end

print("You have successfully exited the query")

break

elif Staff\_count!= 0: #If there are more than 0 matches on the full name of the staff, this will be printed.

print('The person you are checking called ' + xyz + ' was/is a staff of the organization.')

elif Staff\_count== 0: #If there are no matches on the full name of the staff

if xyz.isalpha(): #If there are no matches on the full name of the staff and the user input are letters, this will be printed.

print('The person you are checking called ' + xyz + ' is not a staff of the organization.')

elif type((ast.literal\_eval(xyz))) is float: #If there are no matches on the full name of the staff and the user input is float, this will be printed.

print('This is a float, please enter the full name of person')

elif type((ast.literal\_eval(xyz))) is int: #If there are no matches on the full name of the staff and the user input is integer, this will be printed.

print('This is an integer, please enter the full name of person')

**Python output:**

**Figure 15**

*Sample user input and python output*

**Text, letter

Description automatically generated**

**Explanation:**

A while-loop was used to allow the user to continue to make multiple queries in this interactive user input. An interactive user input was developed by using input() function and what the user keyed in is stored in variable ‘xyz’.

For the variable Staff\_count, the python will search column ‘Staff’ to find a full match of the name in ‘xyz’, it will return a value of more than 0 if it is a full match (value depends on how many times it was matched) and a value of 0 if there was no match.

In the first level of the if-conditional statement, under the if-condition, an operator ‘==’ and break command was used. The code will check if the user entered “exit” which is stored in ‘xyz’. If so, the loop will be broken and a string "You have successfully exited the query" will be printed as output. In the first level of the if-conditional statement, under the first elif-condition, the code will check the variable Staff\_count using the operator ‘!=’ against 0. If there are more than 0 matches on the full name of the staff, a string using String concatenation of ‘+’ operator “The person you are checking called ' + xyz + ' was/is a staff of the organization.” will be printed. In the first level of the if-conditional statement, under the second elif-condition, the code will check the variable Staff\_count using the operator ‘==’ against 0. If there are 0 matches on the full name of the staff, the second level of if-conditional statement will be applied.

In the second level of the if-conditional statement, under the if-condition, isalpha() function was used. If there are 0 matches on the full name of the staff, and letters were inputted by the user, a string using String concatenation of ‘+’ operator “The person you are checking called ' + xyz + ' is not a staff of the organization.” will be printed. In the second level of the if-conditional statement, under the elif-condition, literal\_eval() function in the “ast” class of built-in class library was used on variable ‘xyz’ with ‘is’ keyword. This will check if the number the user inputted was a float type. If so, a string “'This is a float, please enter the full name of person.” will be printed. The next elif will check if the user inputted an integer type. If so, a string “'This is an integer, please enter the full name of person.” will be printed.