HR ANALYTICS - ABSENTEEISM

DETAILED PROJECT REPORT

BY:

SHREEDEVI R

PROJECT DETAIL

Project Title	HR Analytics – Absenteeism Dataset
Technologies	Business Intelligence
Domain	HR
Project Difficulties level	Advanced

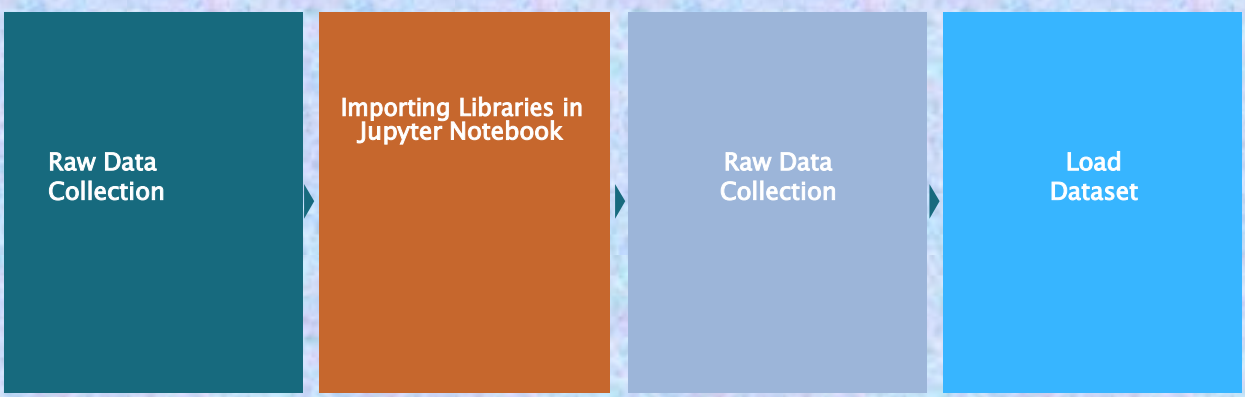
OBJECTIVE

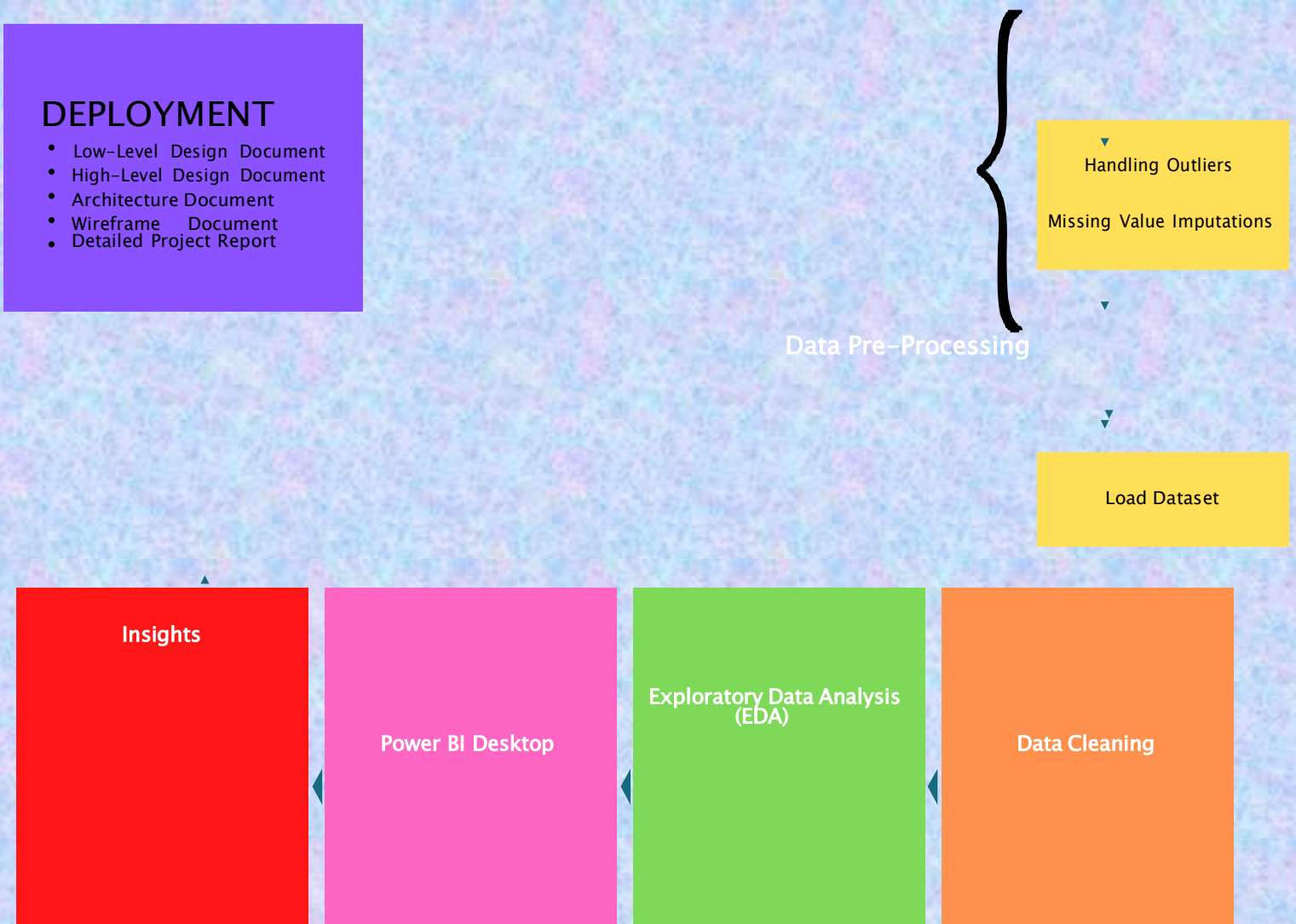
- This data set is suitable for identifying pockets of absence in the organization. These pockets may require interventions. 'Absent Hour' will be used as a dependent variable. How age and length of service may also be associated with absence

PROBLEM STATEMENT

- HR is not just about hiring people it is an ocean of its own. HR department goes through a constant journey of finding, selecting, onboarding and monitoring the right talent. You are required to use analytics concept to provide a smooth monitoring of workforce for the HR department.
- The data set can also be used as an exercise set to predict absence using decision trees or linear models. This data set is quite straightforward. It is large but still manageable in software like SPSS or Excel.
- You may have to code a number of nominal variables into number values before you can do your analysis but on top of that, the data itself doesn't pose much of a challenge. This HR data set focuses on absence at work. The data set contains 740 rows and 21columns of data.
- The data set contains a number of employee IDs. Each row represents a certain quantity of absence – meaning that one employee can have multiple rows.
- Information on employees include the number of children, work load, distance from work, transportation expense, education, height, weight, BMI, and absenteeism time in hours. Other information includes the season, month of absence, day of absence, and day of the week.
- This data set can help you find predictors of absence. Potential analyses could be to see if there is an association between BMI and absence, as well as season, work load, distance from work and the other factors in the data set.

ARCHITECTURE





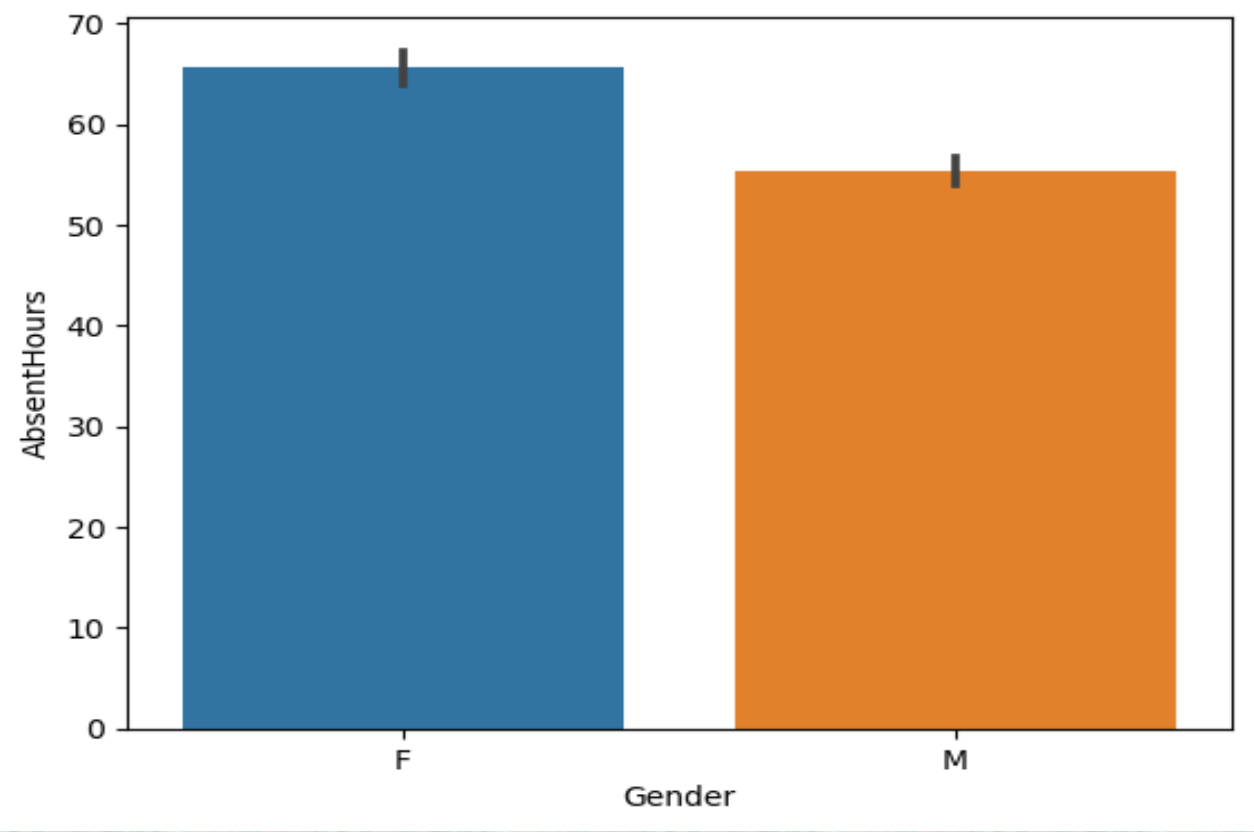
DATASET INFORMATION

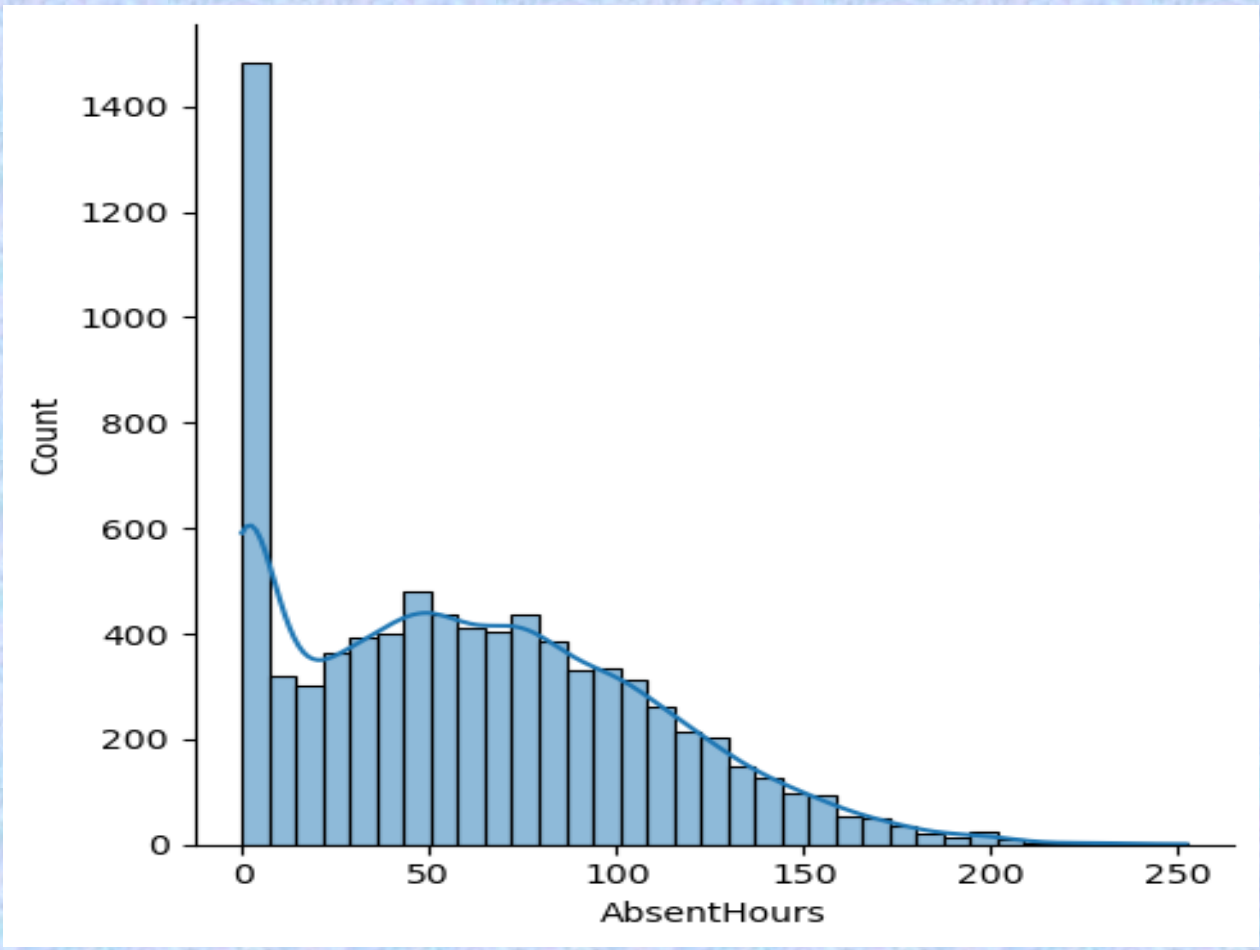
- This data set is neatly structured. This means that every employee has a single line and
  that absence is taken as the total annual absent hours per employee.
- Datasets is available in the given link. This enormous HR data set focuses on employee absence.
- It contains a staggering 8335 rows and 13 columns of data.
- The data set contains employee numbers and names, gender, city, job title, department, store location, business unit, division, age, length of service, and the number of hour absent. You can download as per your convenient.

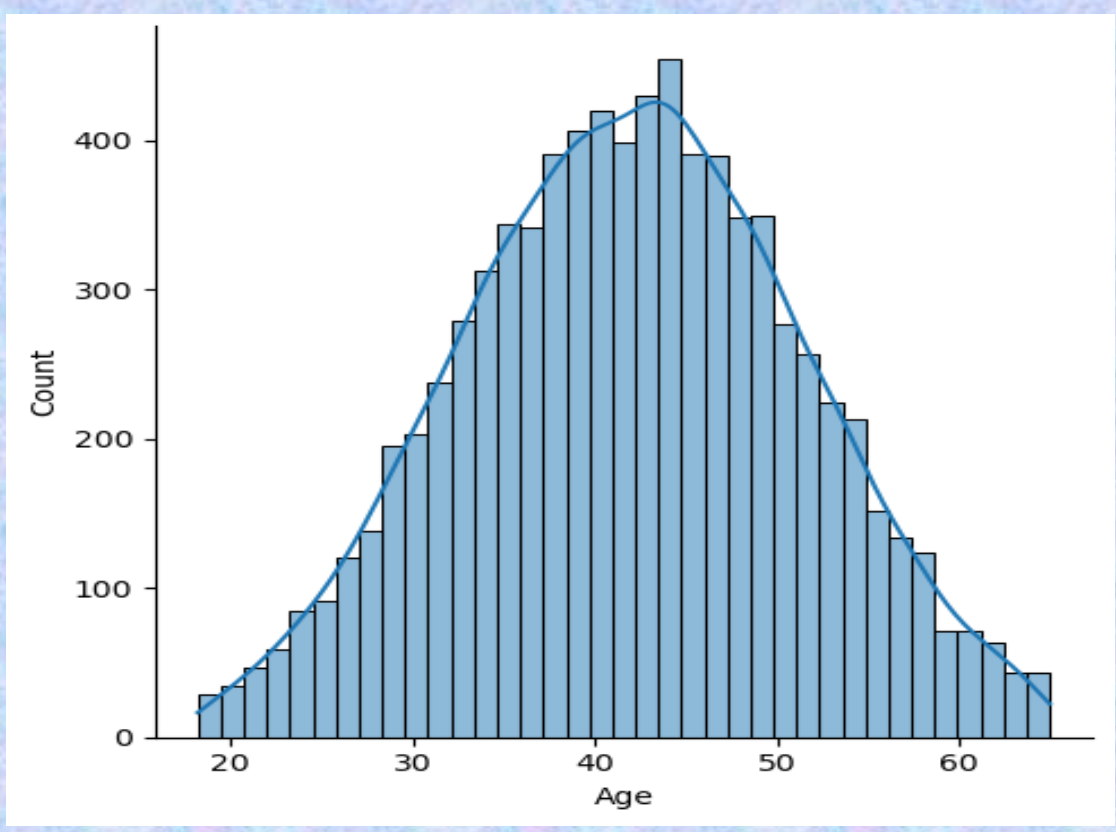
Dataset:

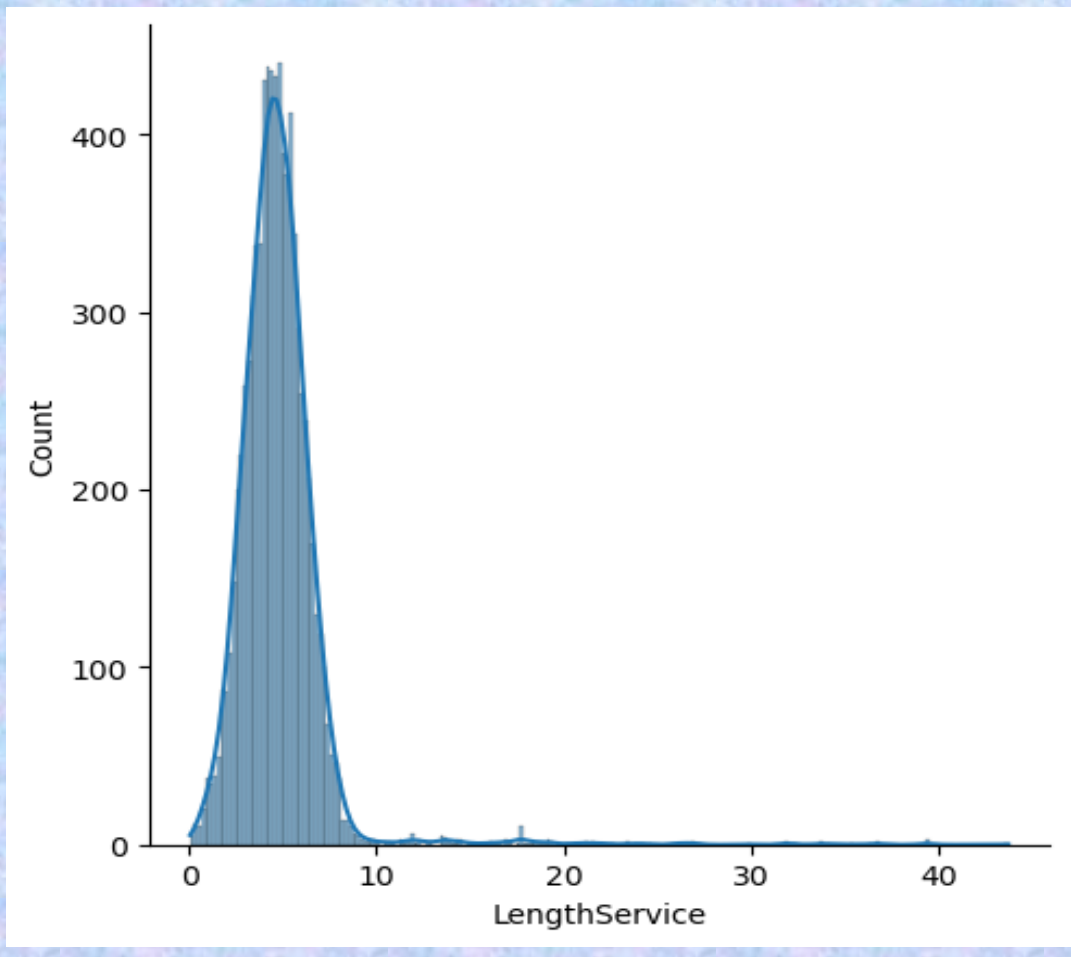
Dataset is available in the given link. You can download as per your convenient.

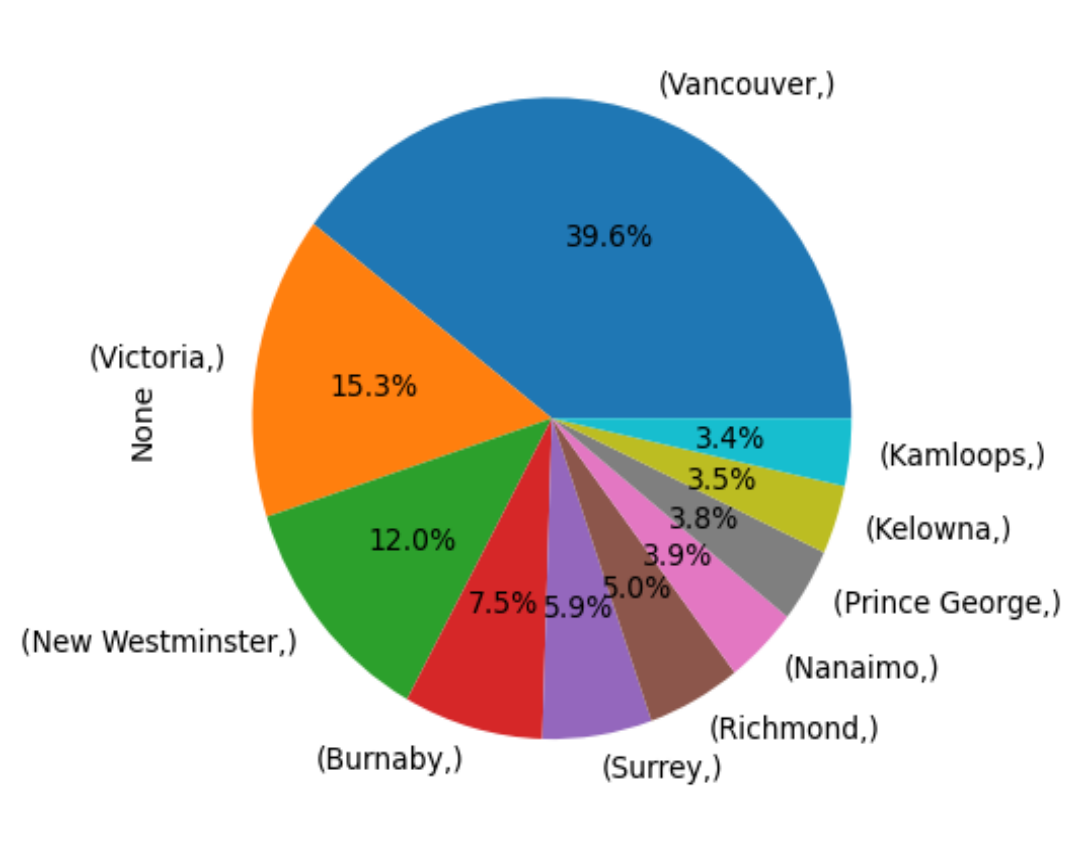
https://drive.google.com/drive/folders/1HRRC23RoCI0MwhV3P_cUz47TNOg4R-_2?usp=sharing

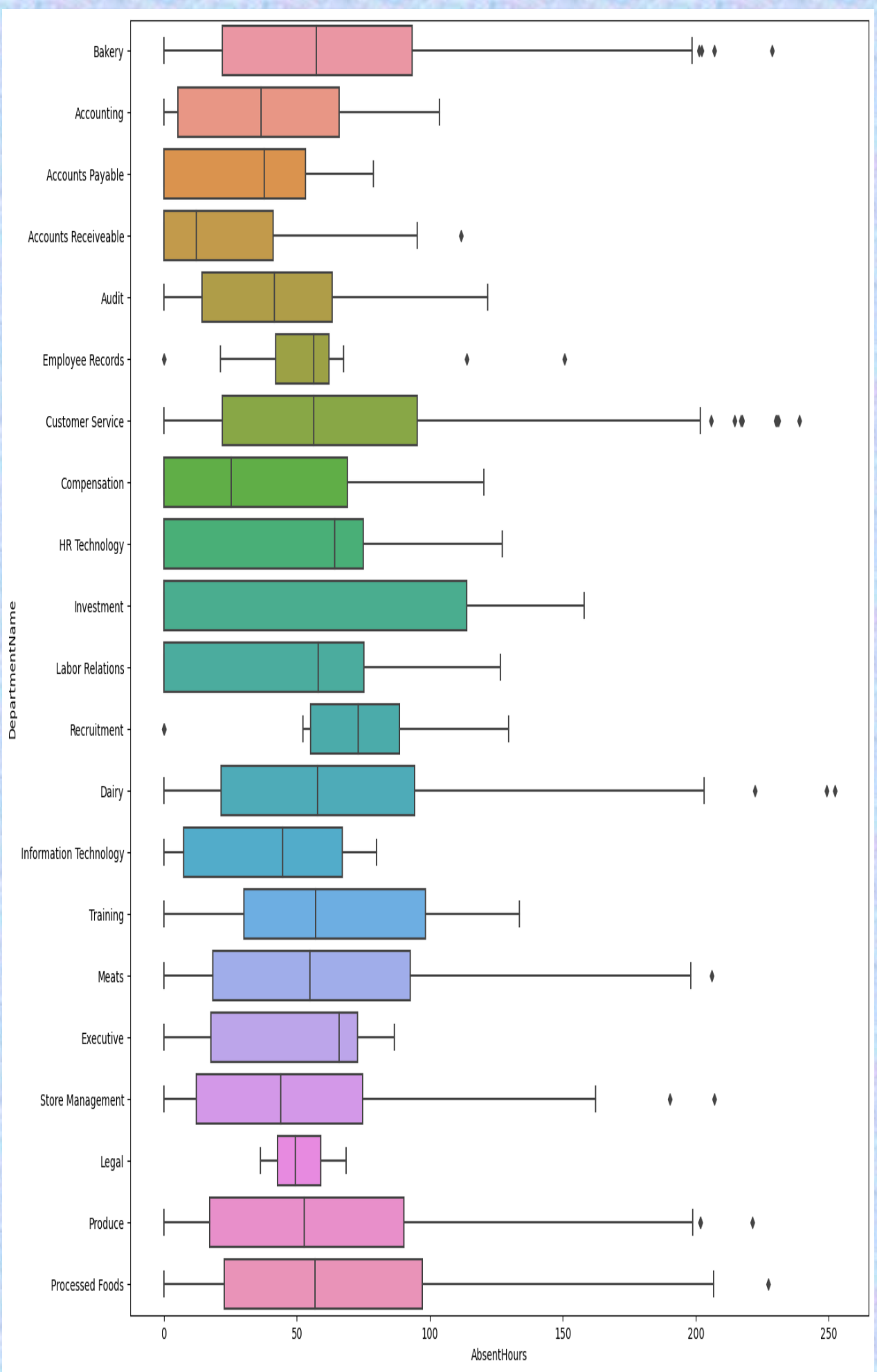


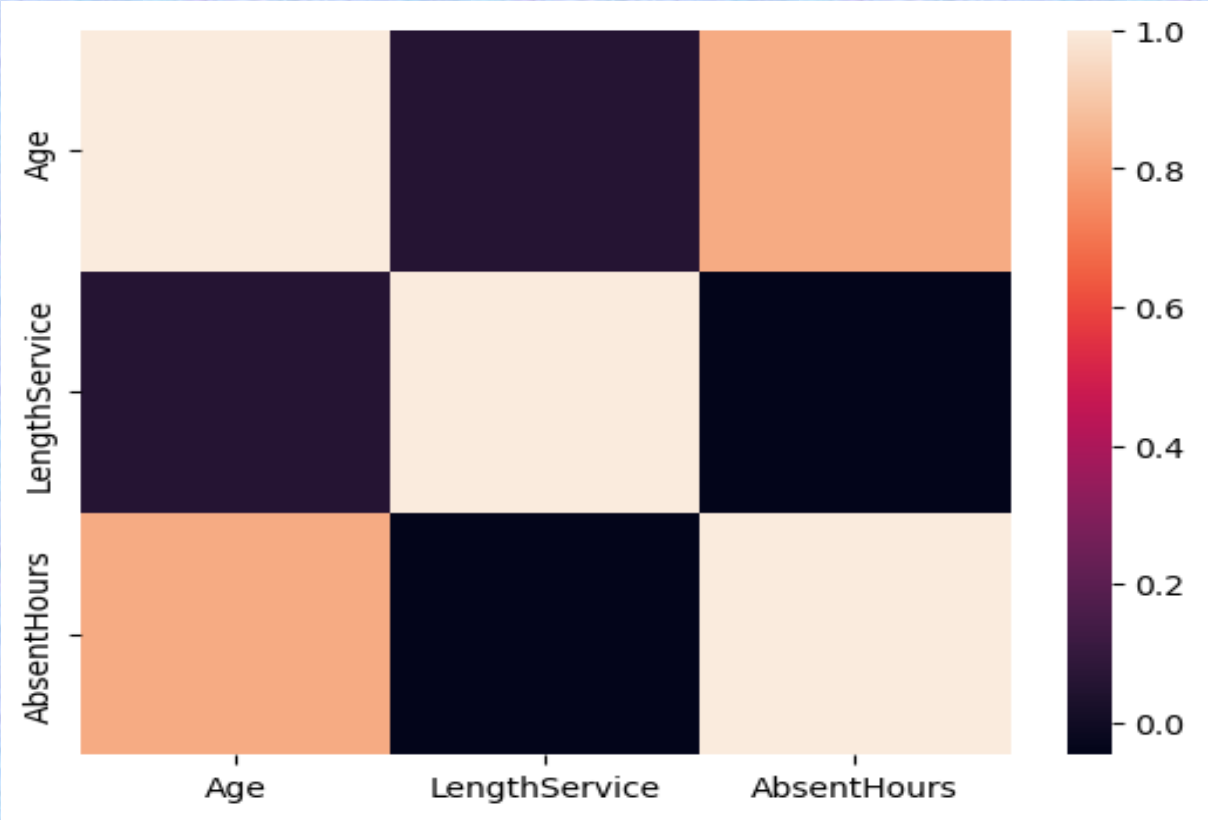

1400
1200
1000
800
600
400
200
0
Count
0
50
100
150
200
250
AbsentHours

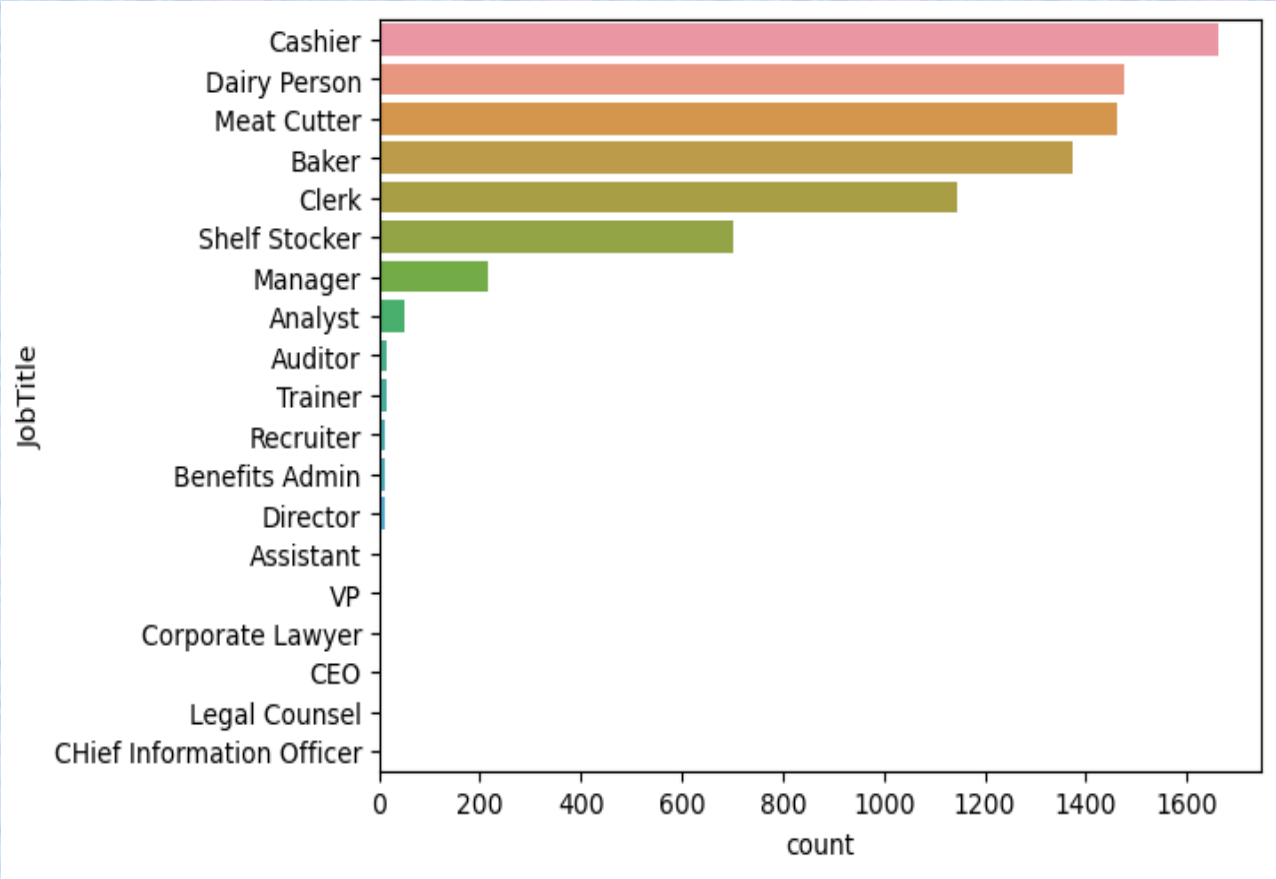

400
300
200
100
0
Count
20
30
40
50
60
Age


Count
400
300
200
100
0
0
10
20
30
40
LengthService

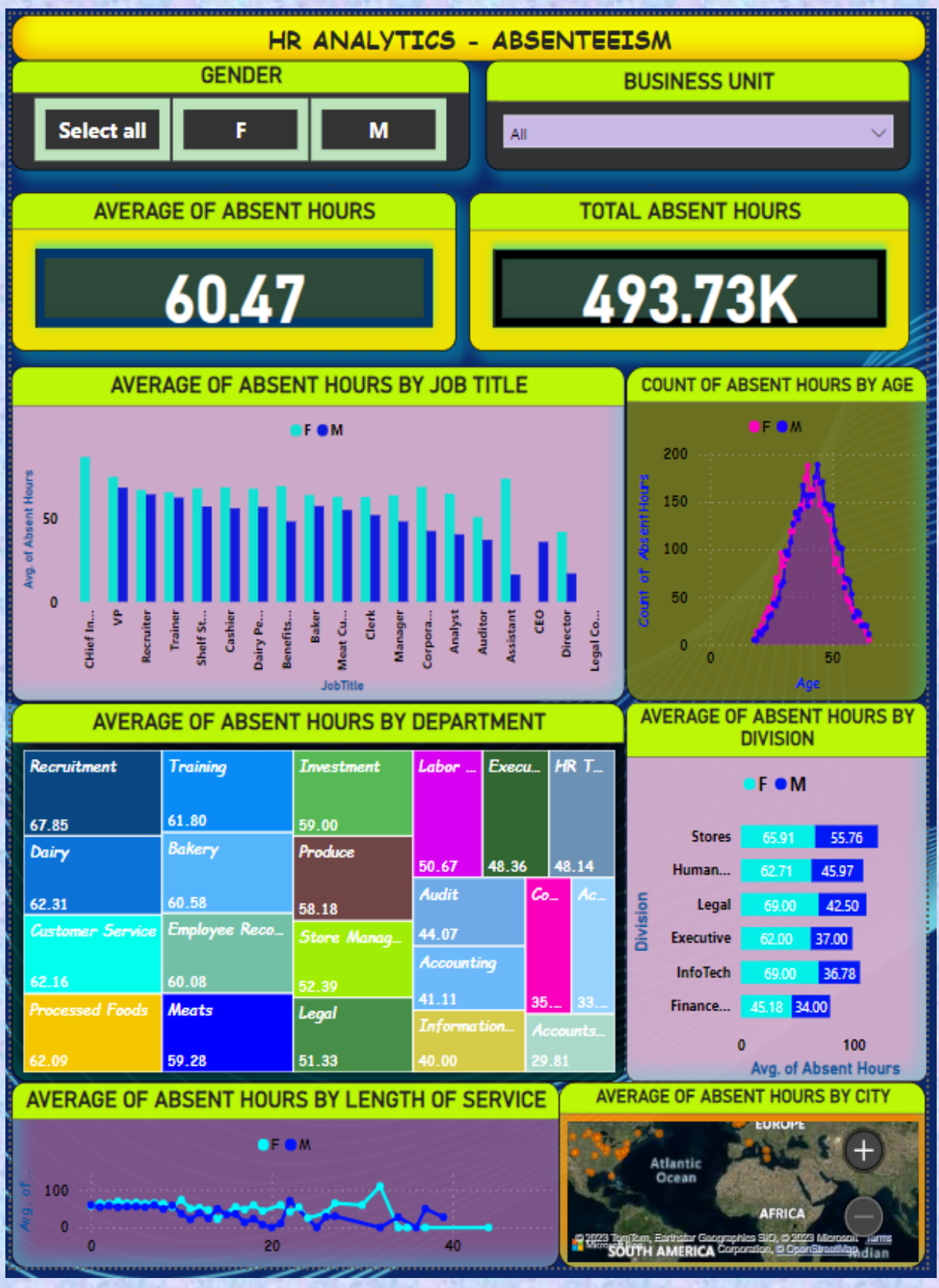

(Vancouver,)
39.6%
(Victoria,)
15.3%
None
3.4%
(Kamloops,)
3.5%
(Kelowna,)
3.8%
3.9%
5.0%
(Prince George,)
12.0%
7.5%
5.9%
(New Westminster,)
(Nanaimo,)
(Richmond,)
(Burnaby,)
(Surrey,)


Bakery
Accounting
Accounts Payable
Accounts Receiveable
Audit
Employee Records
Customer Service
Compensation
HR Technology
Investment
Labor Relations
Recruitment
Dairy
Information Technology
Training
Meats
Executive
Store Management
Legal
Produce
Processed Foods
DepartmentName
0
50
100
150
200
250
AbsentHours


Age
LengthService
AbsentHours
1.0
0.8
0.6
0.4
0.2
0.0
Age
LengthService
AbsentHours


JobTitle
Cashier
Dairy Person
Meat Cutter
Baker
Clerk
Shelf Stocker
Manager
Analyst
Auditor
Trainer
Recruiter
Benefits Admin
Director
Assistant
VP
Corporate Lawyer
CEO
Legal Counsel
CHief Information Officer
0
200
400
600
800
1000
1200
1400
1600
count

• DASHBOARD:


HR ANALYTICS - ABSENTEEISM
GENDER
Select all
F
M
BUSINESS UNIT
All
AVERAGE OF ABSENT HOURS
60.47
TOTAL ABSENT HOURS
493.73K
AVERAGE OF ABSENT HOURS BY JOB TITLE
F M
Avg. of Absent Hours
50
0
Chief In...
VP
Recruiter
Trainer
Shelf St...
Cashier
Dairy Pe...
Benefits...
Baker
Meat Cu...
Clerk
Manager
Corpora...
Analyst
Auditor
Assistant
CEO
Director
Legal Co...
JobTitle
COUNT OF ABSENT HOURS BY AGE
F M
200
150
100
50
0
Count of Absent Hours
0
50
Age
AVERAGE OF ABSENT HOURS BY DEPARTMENT
Recruitment 67.85
Training 61.80
Investment 59.00
Labor ... 50.67
Execu... 48.36
HR T... 48.14
Dairy 62.31
Bakery 60.58
Produce 58.18
Audit 44.07
Co... 35...
Ac... 33...
Customer Service 62.16
Employee Reco... 60.08
Store Manag... 52.39
Accounting 41.11
Processed Foods 62.09
Meats 59.28
Legal 51.33
Information... 40.00
Accounts... 29.81
AVERAGE OF ABSENT HOURS BY DIVISION
F M
Division
Stores 65.91 55.76
Human... 62.71 45.97
Legal 69.00 42.50
Executive 62.00 37.00
InfoTech 69.00 36.78
Finance... 45.18 34.00
0
100
Avg. of Absent Hours
AVERAGE OF ABSENT HOURS BY LENGTH OF SERVICE
F M
Avg. of...
100
0
0
20
40
AVERAGE OF ABSENT HOURS BY CITY
EUROPE
Atlantic Ocean
AFRICA
SOUTH AMERICA
Indian

FINAL INSIGHTS: -

1. The ratio of men and women working are almost equal.
2. The employees working in Stores are more as compared to Head office.
3. Most of the stores are present in Vancouver, Victoria, Nanaimo, New Westminster in the following order.
4. Most of the people are working as Cashier then as Dairy Persons
5. Age and Absent Hours are highly correlated. we can say higher the age higher will be the absent hours.
6. Scatterplot shows the relations between Age and Absent Hours as age increases, Absent Hours also increases.
7. From the Bar plot between Gender and Absent Hours we can say that female ought to have more absent hours.
8. And from Boxplot we can say that females have more outliers.
9. As we can see from the boxplot between Absent Hours and Store Location Vancouver has the most outliers than any store location which means, many employees with high number of absent hours are from Vancouver.

10. We can see employees who were working as Cashiers were comparatively Absent more.
11. Customer Service' and 'Bakery' have the most outliers i.e. Employee from these departments have the Most Absent Hours.
12. Age is Normally Distributed.
13. Length Service is Slightly Skewed Right.
14. Absent Hours is Log Normally Distributed i.e. skewed towards right which means: There are very few employees which are absent for longer period of time and vice versa...

Limitations and Conclusions: -

- The data provided do not include the calendar so we can't draw a conclusion on whether the weather affects the presence of employees at work.
- Also, we can't tell for sure the mode of each worker, if it is remote or otherwise. This information will be needed for further analysis to be carried out. The higher the number of employees in a department and higher the absent hours/days.

THANK YOU