Numerical Dataset

****General Information :**

The Name of Dataset : Salary Data Prediction

That predict the Salary upon some features

Features : ["Age", "Gender", "Education Level", "Job Title", "Years of Experience"]

Label:["Salary"]

Total Number of Samples : 2552

Number of Training : 2042

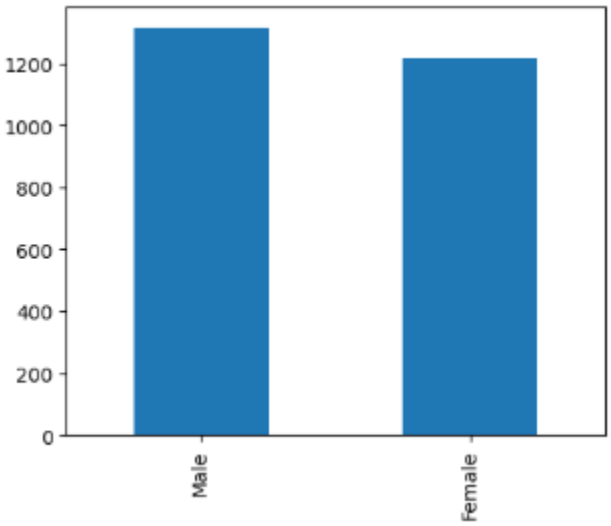
Number of Testing : 510

Some Visualization about Dataset:

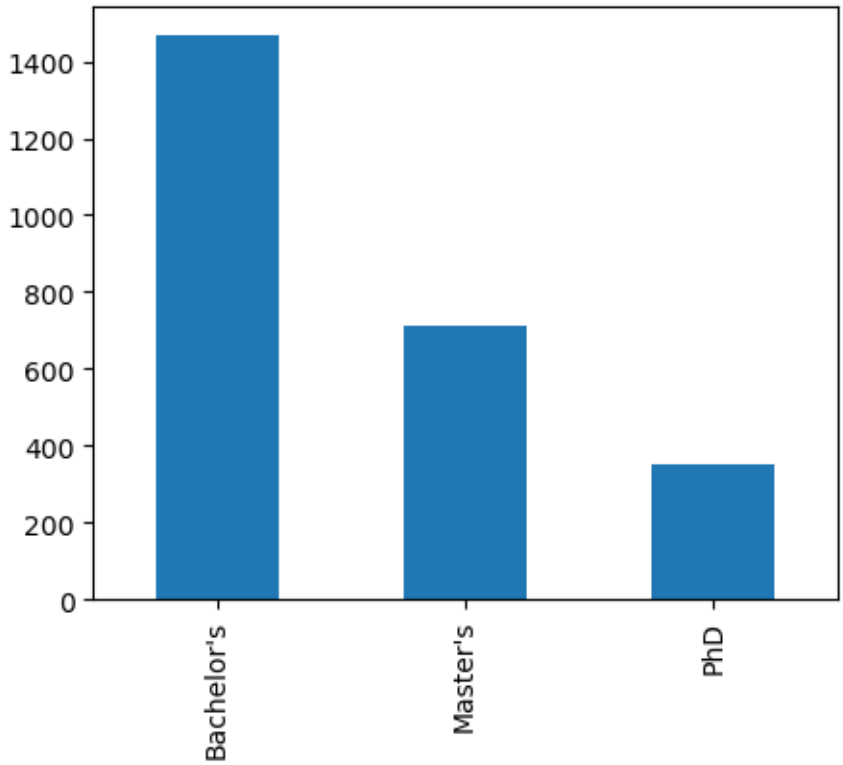
Age vs Salary

Years of Experience vs Salary

Occurrence of Gender values



Occurrence of Education Level values



Max, Min and Average of Salaries

Implementation Details for Linear Regression:

Number of Training : 2042

Number of Testing : 510

Linear Regression Model

```
from sklearn.model_selection import KFold
from sklearn.linear_model import LinearRegression
from sklearn.metrics import accuracy_score ,r2_score

linear_model = LinearRegression()
linear_model.fit(X_train, y_train)

y_pred = linear_model.predict(X_test)

r2 = r2_score(y_test, y_pred)
```

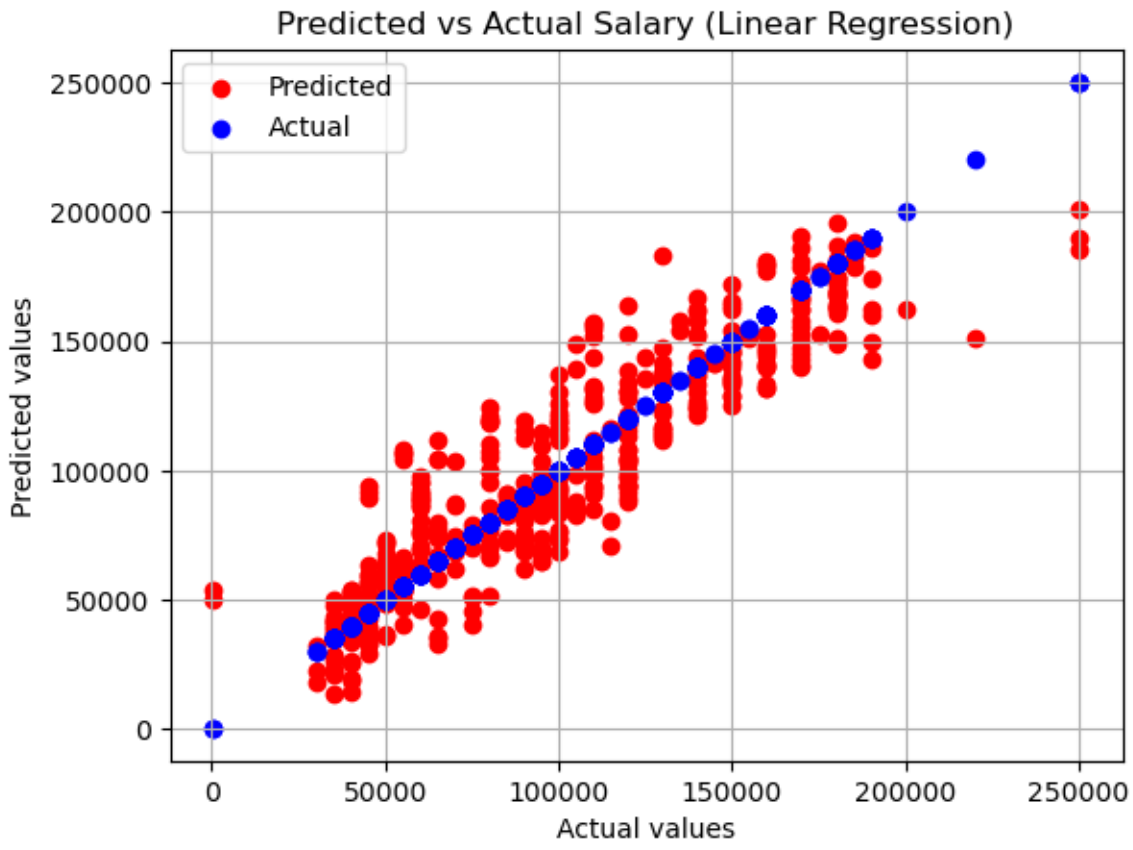
Final Result :

Accuracy:

```
from sklearn.metrics import accuracy_score ,r2_score
r2 = r2_score(y_test, y_pred)
print("Average R2 Score is {0:.2f}% :". format( 100*r2))
```

```
Average R2 Score is 84.10% :
```

Predicted vs Actual Salary :



Implementation Details for KNN Regression:

Number of Training : 2042

Number of Testing : 510

KNN Model

```
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import accuracy_score ,r2_score

knn_regressor = KNeighborsRegressor(n_neighbors=6)
knn_regressor.fit(X_train, y_train)
y_pred = knn_regressor.predict(X_test)
r2 = r2_score(y_test, y_pred)
```

Final Results:

Accuracy:

```
from sklearn.metrics import accuracy_score ,r2_score
print("Best R2 Score is {0:.2f}% :". format( 100*r2))
```

```
Best R2 Score is 86.02% :
```

Predicted vs Actual Salary :