The preprocessing techniques:

1. Column (Date) that has datetime64 as data type needs to be convert into its columns as: day, months ,year (numerical or categorical datatype) using : (to_datetime)
2. Columns (Snowfall,Rainfall) have highly skewed that replace the minimum value with mean or median value
3. Convert the categorical columns (Season ,holiday ,functioning day) to numerical columns using :Label encoder

The Analysis Process :

1. Using function info () to provide information about the data set over all
2. Using function describe () to provide the numerical distribution of numerical columns
3. Using seaborn and matplotlib library to visualize the correlation between The columns, between the target column and other columns
4. Making some analysis and visualizations to recognize the data set and the relations between the features
5. Remove columns(Date, Dew point temperature(°C),Year) That have negative effect on the data

The Regression Techniques:

1. **Linear Regression**: simple linear model used in the field of predictive.
   - using Lasso and ridge regression as perform regularization on the linear regression model .
   - the accuracy of the model is less compared to the other models
   - the error (root mean squared error) is :410

2. **Random Forest**: algorithm combines ensemble learning methods with the decision tree framework to create multiple randomly drawn decision trees from the data.
   - It gives high accuracy compared to the other models
   - the error (root mean squared error) is : 201

3. **Extra Tree Regressor** : algorithm combines ensemble learning methods like random forest but differs from it in :

- Random forest uses bootstrap replicas but Extra Trees use the whole original sample
- Random Forest chooses the optimum split while Extra Trees chooses it randomly
- In terms of computational cost, and therefore execution time, the Extra Trees algorithm is faster
- **Extra Tree** gives high accuracy compared to the other models
- the error (root mean squared error) is : 291

The Features (used or discarded) :

The features used are: (Hour , Temperature(Â°C), Humidity(%), Wind speed (m/s), Visibility (10m), Solar Radiation (MJ/m2), Rainfall(mm), Snowfall (cm), Seasons, Holiday, Functioning Day, Months, Weekday ,Label_day_neight).

The features discarded are : (Date , Dew point temperature(Â°C),Year, Rented Bike Count)

The Size : the size of training data is 80 % and the size of test data is 20 % and using the RANDOM_SEED

the seed value :is used to generate the random number generator and every time you use the same seed value, you will get the same random values.

improve the results :

using the **Grid Search** technique for model tuning : is used to find the optimal hyperparameters of a model which results in the most accurate predictions.

The conclusion:

This phase In short, it contains a set of characteristics used to predict and to know the rented bike counts by understanding these characteristics and their relationship to each other and the ability to build a model capable of predicting with the lowest error rate.

The problem was the high error in results and was solved by make the suitable preprocessing techniques and better models with this data set