```
In [1]: # Import necessary libraries
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        # Load dataset
        df = pd.read csv("HR Analytics.csv")
        # Basic overview
        print(df.shape)
        print(df.info())
        print(df.describe())
        print(df['Attrition'].value_counts())
        # Convert 'Attrition' to binary
        df['Attrition'] = df['Attrition'].map({'Yes': 1, 'No': 0})
        # Attrition count
        plt.figure(figsize=(5, 4))
        sns.countplot(x='Attrition', data=df)
        plt.title('Attrition Count')
        plt.xticks([0,1], ['No', 'Yes'])
        plt.show()
        # Attrition by Department
        plt.figure(figsize=(6, 4))
        sns.countplot(x='Department', hue='Attrition', data=df)
        plt.title('Attrition by Department')
        plt.xticks(rotation=20)
        plt.show()
        # Attrition by Age
        plt.figure(figsize=(6, 4))
        sns.histplot(data=df, x='Age', hue='Attrition', multiple='stack', bins=20)
        plt.title('Attrition by Age')
        plt.show()
        # Attrition by Years at Company
        plt.figure(figsize=(6, 4))
        sns.boxplot(x='Attrition', y='YearsAtCompany', data=df)
        plt.title('Attrition by Years at Company')
        plt.show()
        # Attrition by Job Role
        plt.figure(figsize=(8, 4))
        sns.countplot(y='JobRole', hue='Attrition', data=df)
        plt.title('Attrition by Job Role')
        plt.show()
        # Attrition by Education Field
        plt.figure(figsize=(8, 4))
        sns.countplot(y='EducationField', hue='Attrition', data=df)
        plt.title('Attrition by Education Field')
        plt.show()
```

> (1480, 38) <class 'pandas.core.frame.DataFrame'> RangeIndex: 1480 entries, 0 to 1479 Data columns (total 38 columns):

#	Column	Non-Null Count	Dtype		
0	EmpID	1480 non-null	object		
1	Age	1480 non-null	int64		
2	AgeGroup	1480 non-null	object		
3	Attrition	1480 non-null	object		
4	BusinessTravel	1480 non-null	object		
5	DailyRate	1480 non-null	int64		
6	Department	1480 non-null	object		
7	DistanceFromHome	1480 non-null	int64		
8	Education	1480 non-null	int64		
9	EducationField	1480 non-null	object		
10	EmployeeCount	1480 non-null	int64		
11	EmployeeNumber	1480 non-null	int64		
12	EnvironmentSatisfaction	1480 non-null	int64		
13	Gender	1480 non-null	object		
14	HourlyRate	1480 non-null	int64		
15	JobInvolvement	1480 non-null	int64		
16	JobLevel	1480 non-null	int64		
17	JobRole	1480 non-null	object		
18	JobSatisfaction	1480 non-null	int64		
19	MaritalStatus	1480 non-null	object		
20	MonthlyIncome	1480 non-null	int64		
21	SalarySlab	1480 non-null	object		
22	MonthlyRate	1480 non-null	int64		
23	NumCompaniesWorked	1480 non-null	int64		
24	Over18	1480 non-null	object		
25	OverTime	1480 non-null	object		
26	PercentSalaryHike	1480 non-null	int64		
27	PerformanceRating	1480 non-null	int64		
28	RelationshipSatisfaction	1480 non-null	int64		
29	StandardHours	1480 non-null	int64		
30	StockOptionLevel	1480 non-null	int64		
31	TotalWorkingYears	1480 non-null	int64		
32	TrainingTimesLastYear	1480 non-null	int64		
33	WorkLifeBalance	1480 non-null	int64		
34	YearsAtCompany	1480 non-null	int64		
35	YearsInCurrentRole	1480 non-null	int64		
36	YearsSinceLastPromotion	1480 non-null	int64		
37	YearsWithCurrManager	1423 non-null	float64		
dtype	es: float64(1), int64(25),	object(12)			
		J \ /			

memory usage: 439.5+ KB

None

std

min 25%

50%

	Age	DailyRate	DistanceFromHom	ne Educatio	n EmployeeCount	\
count	1480.000000	1480.000000	1480.00000	0 1480.00000	0 1480.0	
mean	36.917568	801.384459	9.22027	0 2.91081	1.0	
std	9.128559	403.126988	8.13120	1.02379	6 0.0	
min	18.000000	102.000000	1.00000	0 1.00000	0 1.0	
25%	30.000000	465.000000	2.00000	0 2.00000	0 1.0	
50%	36.000000	800.000000	7.00000	3.00000	0 1.0	
75%	43.000000	1157.000000	14.00000	0 4.00000	0 1.0	
max	60.000000	1499.000000	29.00000	0 5.00000	0 1.0	
	EmployeeNumb	er Environme	ntSatisfaction	HourlyRate	JobInvolvement	\
count	1480.0000	100	1480.000000	1480.000000	1480.000000	
mean	1031.8608	11	2.724324	65.845270	2.729730	

1.092579

1.000000

2.000000

3.000000

20.328266

30.000000

48.000000

66.000000

605.955046

493.750000

1027.500000

1.000000

0.713007

1.000000

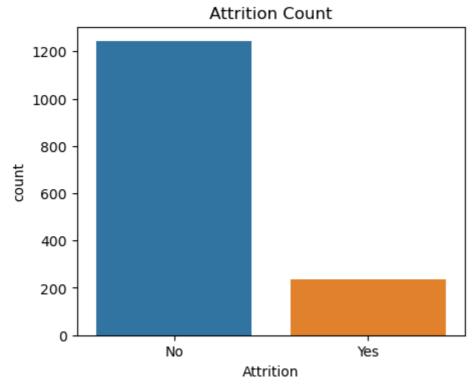
2.000000

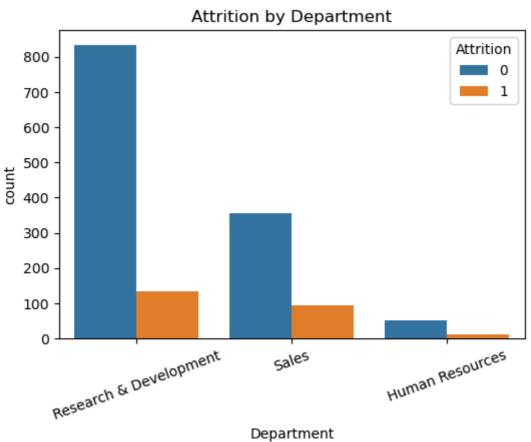
3.000000

```
75%
          1568.250000
                                        4.000000
                                                    83.000000
                                                                       3.000000
          2068.000000
                                                   100.000000
                                                                       4.000000
                                        4.000000
max
          JobLevel
                          RelationshipSatisfaction StandardHours
count 1480.000000
                                        1480.000000
                                                             1480.0
mean
          2.064865
                                           2.708784
                                                               80.0
          1.105574
                                           1.081995
                                                                0.0
std
                     . . .
min
          1.000000
                                           1.000000
                                                               80.0
25%
          1.000000
                                           2.000000
                                                               80.0
                     . . .
50%
          2.000000
                                           3.000000
                                                               80.0
75%
          3.000000
                                           4.000000
                                                               80.0
max
          5.000000
                                           4.000000
                                                               80.0
                         TotalWorkingYears
       StockOptionLevel
                                              TrainingTimesLastYear
            1480.000000
                                1480.000000
                                                         1480.000000
count
mean
                0.791892
                                   11.281757
                                                            2.797973
std
                0.850527
                                    7.770870
                                                            1.288791
min
                0.000000
                                    0.000000
                                                            0.000000
25%
                0.000000
                                    6.000000
                                                            2.000000
50%
                1.000000
                                   10.000000
                                                            3.000000
75%
                1.000000
                                   15.000000
                                                            3.000000
                3.000000
                                   40.000000
                                                            6.000000
max
       WorkLifeBalance YearsAtCompany
                                          YearsInCurrentRole
           1480.000000
                            1480.000000
                                                 1480.000000
count
              2.760811
                               7.009459
                                                    4.228378
mean
              0.707024
std
                               6.117945
                                                     3.616020
              1.000000
                               0.000000
                                                    0.000000
min
25%
              2.000000
                               3.000000
                                                     2.000000
50%
              3.000000
                               5.000000
                                                     3.000000
75%
                               9.000000
              3.000000
                                                    7.000000
max
              4.000000
                               40.000000
                                                    18.000000
       YearsSinceLastPromotion YearsWithCurrManager
count
                    1480.000000
                                           1423.000000
mean
                       2.182432
                                              4.118060
std
                       3.219357
                                              3.555484
min
                       0.000000
                                              0.000000
25%
                       0.000000
                                              2.000000
50%
                       1.000000
                                              3.000000
75%
                       3.000000
                                              7.000000
                      15.000000
                                             17.000000
max
[8 rows x 26 columns]
Attrition
No
       1242
```

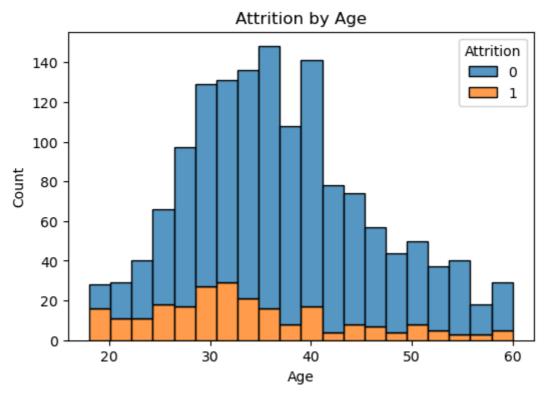
238 Yes

Name: count, dtype: int64

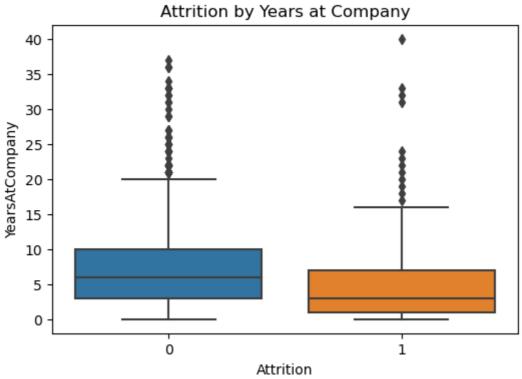


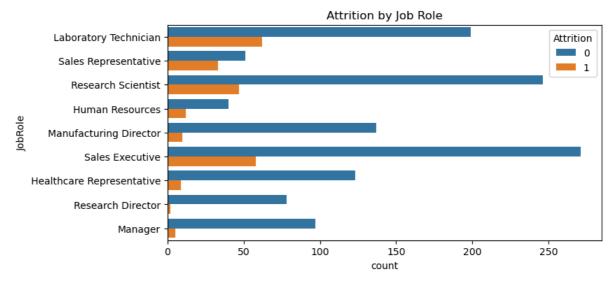


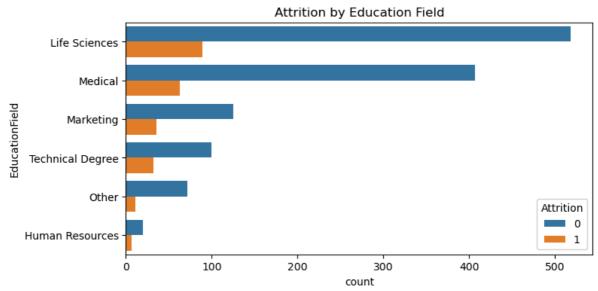
5/7/25, 5:05 PM



HR EDA







EmpID 0 0 Age AgeGroup 0 Attrition 0 BusinessTravel 0 DailyRate 0 Department 0 DistanceFromHome 0 Education 0 EducationField 0 0 EmployeeCount EmployeeNumber EnvironmentSatisfaction 0 Gender 0 HourlyRate 0 JobInvolvement 0 JobLevel 0 JobRole 0 JobSatisfaction 0 MaritalStatus 0 MonthlyIncome 0 SalarySlab 0 MonthlyRate 0 NumCompaniesWorked 0 Over18 0 OverTime 0 PercentSalaryHike 0 PerformanceRating 0 RelationshipSatisfaction 0 StandardHours StockOptionLevel 0 TotalWorkingYears 0 TrainingTimesLastYear WorkLifeBalance 0 YearsAtCompany 0 YearsInCurrentRole 0 YearsSinceLastPromotion 0 YearsWithCurrManager 57

dtype: int64

```
In [9]: plt.figure(figsize=(14,10))
    sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
    plt.title('Feature Correlation Heatmap')
    plt.show()
```

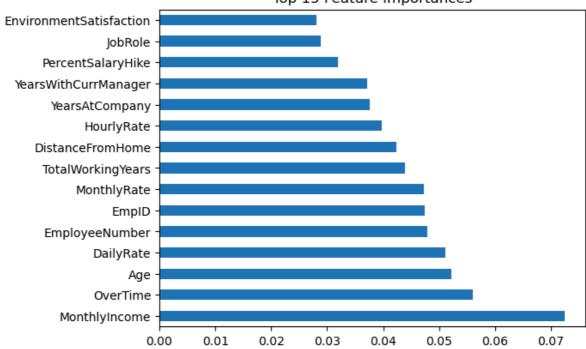
```
EmpID - 1,000 880.63200 90 850260630 490 17 -0.0670 48.0560 650 4010 280 36.0010 430 430 200 07 0.06.0280 850 29 -0.00 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020 16.0020
                                                                                      81 0.9<mark>5</mark>0.1660 260-1220 290 032 20.041 - 0.00.D-1220 8-0 230 27.5-D-1020 635 89 8.50.40.0 30.3
                                                                                                                                                                                                                          0.029008601654 0.03960.00702<mark>230.20.20.21</mark>
                                                            AgeGroup0-002<mark>0.95 1</mark> 0.050260350670109149.04-0.009997.504302031<mark>7.4</mark>90.31.01-2.10.45 0.0107.31 0.0020018905035 0.04<mark>0.50</mark>.02604290.20.210.2
Attrition 9.020.1-6.11 1 0.060.05065030.03025-0.01-2.0060083-5.070680.00.140.161.020.07045 0.26.0316025045 -0.1-4.107.0580631-9.36030216
                                                    BusinessTravel-6.0002602603 1 0067611922001213 -0.027962045050129802190050219 0.0299292049933-0.0004910136116091692604922
                                                                                                                                                                                                                                                                                                                               0.75
                                                        DistanceFromHome0-900.600-03300968.602 00701 1 0.020022 0.0380-0400-402-80634074401694 0-00-00-00-050502 0.0260 0702-8028 0.04-097.80-03101.0 0-09 02011
                                                           Education-0.04921.19.03041.81300697 10.04 0.03809701.010904309.006416054920952713 -0.019.01.0200580.017.16.02608060198012069
                                                    - 0.50
                                  EmployeeNumber-e.057c1.0c09002027c5b007078c080104 1.021025c3.0e7.010600010401.00000530.0260020202074.0.05800.002009201.0e8701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.0e98701.
                                                  0.25
                                                   MaritalStatus -0.01.098.10.1602806995501605.416-0.0009824990170890796822 10.00709828038 -0.0201.404.023
                                                   MonthlyIncome 6.04<mark>3.50.47</mark>0.050 % 10 და 940 და 94
                                                                                                                                                                                                                                                                                                                              - 0.00
                                                           SalarySlab-0.043
                                                       MonthlyRate-9.02903.0 D30.D6030250260027026 0.003964.042013902003.9 053001.0 203-03405 1 0.02 0.002.0 0.02.0 0.0322604399.D2500360439
                                     NumCompaniesWorked 6,00<mark>7,30.31</mark>,0450 D93904570 Z91300960.00503103045.0510 1614.0590550 381-5.10.02 1 -0.009901004053 0.03226.06506802.091037.11
                                                                Over18
                                                            -0.25
                                            StandardHours
                                               StockOptionLevel-6.00@199046.143.00.0-4220@30450-197018 0.05.802201305.092016.022019.006563.03.20 0.001007896.9946 1.012201004.905031016.02
                                                                                                                                                                                                                                                                                                                                 -0.50
                                               TotalWorkingYears 9,01<mark>0,68.65</mark>0.1070 410 105011.90 10318-6.029 - 0.01.1902 10318-00 1070.1040 1090 10770 10.026 24 0.0130020 05 12 4 0.01
                                       TrainingTimesLastYear -0.6020070280580.804936096026051 0.022007098001000809.602003694963-0.60206070190190.01203
                                                WorkLifeBalance 9.0-050920440-93000889250000389410.0039229012044903389250200081393.0269900580.0900900580.0900002034180.094490052
                                                YearsAtCompany 0.0 0.310.240.0400160402400.066019-0-01000032810-0002554.0800-0006.50.40.02512-0.040305.00318 0.01005002209 1.0
                                             YearsInCurrentRole 9,0102210,20,00000800,0570 D9958011-0,00000000000999,9020001,00536,3050003091-0,000000000404017 0,051405000404076 10,55
                                   YearsSinceLastPromotion 9,02<mark>8,2,0,2</mark>0,0B209592,920,0D205200680,0020-D409599602<mark>0536</mark>,0B509532235,B60904937-0,0D2095057031_0,01640100220222850055_1_0.5
                                       YearsWithCurrManager 9.01<u>08 210.20,106 09.30 20 550 D136590 420,0090903.192.5003238 30,000 610 105659,03654,302</u>0 308 11 - 0,002006602.0044 0.00<mark>7.46</mark>,900.00 0
                                                                                                                                                                                                MonthlyIncome -
SalarySlab -
MonthlyRate -
NumCompaniesWorked -
Over18 -
                                                                                                                                                        Gender -
HourlyRate -
Joblnvolvement -
JobLevel -
                                                                                                                                                                               JobRole JobSatisfaction
                                                                                                           DailyRate
Department
                                                                                                                             Education
In [12]: from sklearn.model_selection import train test split
                                   X = df.drop('Attrition', axis=1)
                                   y = df['Attrition']
                                  X train, X test, y train, y test = train test split(X, y, test size=0.2, random sta
In [13]: print(X_train.isnull().sum().sort_values(ascending=False).head(10))
                                  YearsWithCurrManager
                                                                                                                                      46
                                  MonthlyIncome
                                  MonthlyRate
                                                                                                                                          0
                                  NumCompaniesWorked
                                                                                                                                          0
                                  Over18
                                                                                                                                          0
                                  OverTime
                                                                                                                                          0
                                  PercentSalaryHike
                                                                                                                                          0
                                  PerformanceRating
                                  RelationshipSatisfaction
                                                                                                                                          0
                                  StandardHours
                                                                                                                                          0
                                  dtype: int64
In [14]: from sklearn.impute import SimpleImputer
                                   imputer = SimpleImputer(strategy='median')
                                   X_train['YearsWithCurrManager'] = imputer.fit_transform(X_train[['YearsWithCurrManager'])
                                   X_test['YearsWithCurrManager'] = imputer.transform(X_test[['YearsWithCurrManager']]
In [15]: from sklearn.linear_model import LogisticRegression
                                   from sklearn.metrics import classification report, confusion matrix, accuracy score
                                   logreg = LogisticRegression(max_iter=1000)
```

Feature Correlation Heatmap

1.00

```
logreg.fit(X_train, y_train)
          y_pred = logreg.predict(X_test)
          print(classification_report(y_test, y_pred))
          print(confusion_matrix(y_test, y_pred))
          print('Accuracy:', accuracy_score(y_test, y_pred))
                                  recall f1-score
                        precision
                                                        support
                     0
                             0.85
                                       1.00
                                                 0.92
                                                            248
                     1
                             0.80
                                       0.08
                                                 0.15
                                                             48
                                                 0.85
                                                            296
             accuracy
                             0.82
                                       0.54
                                                 0.53
                                                            296
            macro avg
         weighted avg
                             0.84
                                       0.85
                                                 0.79
                                                            296
         [[247
                 1]
          [ 44
                 4]]
         Accuracy: 0.847972972973
In [16]: from sklearn.ensemble import RandomForestClassifier
          rfc = RandomForestClassifier(random_state=42)
         rfc.fit(X_train, y_train)
         y_pred_rfc = rfc.predict(X_test)
          print(classification_report(y_test, y_pred_rfc))
          print(confusion_matrix(y_test, y_pred_rfc))
          print('Accuracy:', accuracy_score(y_test, y_pred_rfc))
                        precision
                                     recall f1-score
                                                        support
                     0
                             0.86
                                       0.99
                                                 0.92
                                                            248
                     1
                             0.75
                                       0.19
                                                 0.30
                                                             48
                                                            296
                                                 0.86
             accuracy
                                       0.59
                                                 0.61
                                                            296
            macro avg
                             0.81
                             0.84
                                       0.86
                                                 0.82
                                                            296
         weighted avg
         [[245
                 3]
          [ 39
                 9]]
         Accuracy: 0.8581081081081081
         importances = pd.Series(rfc.feature importances , index=X.columns)
In [17]:
          importances.nlargest(15).plot(kind='barh')
          plt.title('Top 15 Feature Importances')
          plt.show()
```

Top 15 Feature Importances



In []: