

```
import pandas as pd
```

```
features = pd.read_csv("dep_big_data.csv")
```

```
features
```

```
↳
```

| | distance | presidents_day | easter | memorial_day | independence_day | labor_day | thanksgiving | winter_holiday | total_dep_delay |
|--------|----------|----------------|--------|--------------|------------------|-----------|--------------|----------------|-----------------|
| 0 | 479 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 51.0 |
| 1 | 1222 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4.0 |
| 2 | 2381 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19.0 |
| 3 | 594 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 88.0 |
| 4 | 1080 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 48.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 177521 | 594 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 31.0 |
| 177522 | 695 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24.0 |
| 177523 | 1182 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8.0 |
| 177524 | 1846 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 27.0 |
| 177525 | 546 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1.0 |

177526 rows × 382 columns

```
# Use numpy to convert to arrays
```

```
import numpy as np
```

```
# Labels are the values we want to predict
```

```
labels = np.array(features['total_dep_delay'])
```

```
# Remove the labels from the features
```

```
# axis 1 refers to the columns
```

```
features = features.drop(['total_dep_delay'], axis = 1)
```

```
# Saving feature names for later use
```

```
feature_list = list(features.columns)
```

```
# Convert to numpy array
```

```
features = np.array(features)
```

```
# Splitting the dataset into the Training set and Test set
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(features, labels, test_size = 0.2, random_state = 42)
```

```
# Fitting Random Forest Regression to the Training set
```

```
from sklearn.ensemble import RandomForestRegressor
```

```
regressor = RandomForestRegressor(n_estimators = 100, random_state = 42)
```

```
regressor.fit(X_train, y_train)
```

```
# Predicting the Test set results
```

```
y_pred = regressor.predict(X_test)
```

```
# Evaluating the Algorithm
```

```
from sklearn import metrics
```

```
print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
```

```
print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
```

```
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
```

```
↳ Mean Absolute Error: 16.458917584544984
Mean Squared Error: 682.3504153722291
Root Mean Squared Error: 26.121837901882575
```

```
# Get numerical feature importances
```

```
importances = list(regressor.feature_importances_)
```

```
# List of tuples with variable and importance
```

```
feature_importances = [(feature, round(importance, 2)) for feature, importance in zip(feature_list, importances)]
```

```
# Sort the feature importances by most important first
```

```
feature_importances = sorted(feature_importances, key = lambda x: x[1], reverse = True)
```

```
# Print out the feature and importances
```

```
[print('Variable: {} Importance: {}'.format(*pair)) for pair in feature_importances];
```

```

Variable: distance      Importance: 0.07
Variable: temp          Importance: 0.07
Variable: dew_point_temp Importance: 0.07
Variable: wind_speed    Importance: 0.07
Variable: sea_level_pressure Importance: 0.06
Variable: station_pressure Importance: 0.06
Variable: visibilty     Importance: 0.04
Variable: prcp          Importance: 0.03
Variable: rain_drizzle  Importance: 0.01
Variable: thunder       Importance: 0.01
Variable: month_3       Importance: 0.01
Variable: month_4       Importance: 0.01
Variable: month_5       Importance: 0.01
Variable: month_6       Importance: 0.01
Variable: day_of_week_1 Importance: 0.01
Variable: day_of_week_2 Importance: 0.01
Variable: day_of_week_3 Importance: 0.01
Variable: day_of_week_4 Importance: 0.01
Variable: day_of_week_5 Importance: 0.01
Variable: day_of_week_6 Importance: 0.01
Variable: day_of_week_7 Importance: 0.01
Variable: American      Importance: 0.01
Variable: JetBlue       Importance: 0.01
Variable: Delta         Importance: 0.01
Variable: Frontier      Importance: 0.01
Variable: Spirit        Importance: 0.01
Variable: United        Importance: 0.01
Variable: Southwest     Importance: 0.01
Variable: origin_JFK    Importance: 0.01
Variable: dest_ATL      Importance: 0.01
Variable: dest_BOS      Importance: 0.01
Variable: dest_DEN      Importance: 0.01
Variable: dest_DFW      Importance: 0.01
Variable: dest_FLL      Importance: 0.01
Variable: dest_JFK      Importance: 0.01
Variable: dest_MCO      Importance: 0.01
Variable: dest_ORD      Importance: 0.01
Variable: presidents_day Importance: 0.0
Variable: easter        Importance: 0.0
Variable: memorial_day  Importance: 0.0
Variable: independence_day Importance: 0.0
Variable: labor_day     Importance: 0.0
Variable: thanksgiving  Importance: 0.0
Variable: winter_holiday Importance: 0.0
Variable: fog           Importance: 0.0
Variable: snow_ice_pellets Importance: 0.0
Variable: hail          Importance: 0.0
Variable: tornado_funnel_cloud Importance: 0.0
Variable: month_1       Importance: 0.0
Variable: month_10      Importance: 0.0
Variable: month_11      Importance: 0.0
Variable: month_12      Importance: 0.0
Variable: month_2       Importance: 0.0
Variable: month_7       Importance: 0.0
Variable: month_8       Importance: 0.0
Variable: month_9       Importance: 0.0
Variable: Alaska        Importance: 0.0
Variable: Allegiant     Importance: 0.0
Variable: Hawaiian      Importance: 0.0
Variable: origin_ABE     Importance: 0.0
Variable: origin_ABQ     Importance: 0.0
Variable: origin_ACY     Importance: 0.0
Variable: origin_ALB     Importance: 0.0
Variable: origin_AMA     Importance: 0.0
Variable: origin_ATL     Importance: 0.0
Variable: origin_AUS     Importance: 0.0
Variable: origin_AVP     Importance: 0.0
Variable: origin_BDL     Importance: 0.0
Variable: origin_BHM     Importance: 0.0
Variable: origin_BIL     Importance: 0.0
Variable: origin_BIS     Importance: 0.0
Variable: origin_BNA     Importance: 0.0
Variable: origin_BOI     Importance: 0.0
Variable: origin_BOS     Importance: 0.0
Variable: origin_BTR     Importance: 0.0
Variable: origin_BTV     Importance: 0.0
Variable: origin_BUF     Importance: 0.0
Variable: origin_BWI     Importance: 0.0
Variable: origin_CAE     Importance: 0.0
Variable: origin_CAK     Importance: 0.0
Variable: origin_CHA     Importance: 0.0
Variable: origin_CHS     Importance: 0.0

```

| | |
|----------------------|-----------------|
| Variable: origin_CLE | Importance: 0.0 |
| Variable: origin_CLT | Importance: 0.0 |
| Variable: origin_CMH | Importance: 0.0 |
| Variable: origin_COS | Importance: 0.0 |
| Variable: origin_CRP | Importance: 0.0 |
| Variable: origin_CRW | Importance: 0.0 |
| Variable: origin_CVG | Importance: 0.0 |
| Variable: origin_DAB | Importance: 0.0 |
| Variable: origin_DAY | Importance: 0.0 |
| Variable: origin_DCA | Importance: 0.0 |
| Variable: origin_DLH | Importance: 0.0 |
| Variable: origin_DSM | Importance: 0.0 |
| Variable: origin_DTW | Importance: 0.0 |
| Variable: origin_ELP | Importance: 0.0 |
| Variable: origin_EUG | Importance: 0.0 |
| Variable: origin_EVV | Importance: 0.0 |
| Variable: origin_EWR | Importance: 0.0 |
| Variable: origin_EYW | Importance: 0.0 |
| Variable: origin_FAR | Importance: 0.0 |
| Variable: origin_FAT | Importance: 0.0 |
| Variable: origin_FCA | Importance: 0.0 |
| Variable: origin_FNT | Importance: 0.0 |
| Variable: origin_FSD | Importance: 0.0 |
| Variable: origin_FWA | Importance: 0.0 |
| Variable: origin_GEG | Importance: 0.0 |
| Variable: origin_GFK | Importance: 0.0 |
| Variable: origin_GJT | Importance: 0.0 |
| Variable: origin_GNV | Importance: 0.0 |
| Variable: origin_GRB | Importance: 0.0 |
| Variable: origin_GRI | Importance: 0.0 |
| Variable: origin_GRR | Importance: 0.0 |
| Variable: origin_GSO | Importance: 0.0 |
| Variable: origin_GTF | Importance: 0.0 |
| Variable: origin_IAD | Importance: 0.0 |
| Variable: origin_IAH | Importance: 0.0 |
| Variable: origin_ILG | Importance: 0.0 |
| Variable: origin_ILM | Importance: 0.0 |
| Variable: origin_IND | Importance: 0.0 |
| Variable: origin_JAX | Importance: 0.0 |
| Variable: origin_LAS | Importance: 0.0 |
| Variable: origin_LAX | Importance: 0.0 |
| Variable: origin_LBB | Importance: 0.0 |
| Variable: origin_LEX | Importance: 0.0 |
| Variable: origin_LGA | Importance: 0.0 |
| Variable: origin_LGB | Importance: 0.0 |
| Variable: origin_LIT | Importance: 0.0 |
| Variable: origin_MAF | Importance: 0.0 |
| Variable: origin_MCO | Importance: 0.0 |
| Variable: origin_MEM | Importance: 0.0 |
| Variable: origin_MFR | Importance: 0.0 |
| Variable: origin_MIA | Importance: 0.0 |
| Variable: origin_MKE | Importance: 0.0 |
| Variable: origin_MLI | Importance: 0.0 |
| Variable: origin_MOB | Importance: 0.0 |
| Variable: origin_MSN | Importance: 0.0 |
| Variable: origin_MSO | Importance: 0.0 |
| Variable: origin_MSP | Importance: 0.0 |
| Variable: origin_MSY | Importance: 0.0 |
| Variable: origin_OKC | Importance: 0.0 |
| Variable: origin_OMA | Importance: 0.0 |
| Variable: origin_ORD | Importance: 0.0 |
| Variable: origin_ORF | Importance: 0.0 |
| Variable: origin_ORH | Importance: 0.0 |
| Variable: origin_PBI | Importance: 0.0 |
| Variable: origin_PDX | Importance: 0.0 |
| Variable: origin_PHL | Importance: 0.0 |
| Variable: origin_PHX | Importance: 0.0 |
| Variable: origin_PIA | Importance: 0.0 |
| Variable: origin_PIT | Importance: 0.0 |
| Variable: origin_PNS | Importance: 0.0 |
| Variable: origin_PVD | Importance: 0.0 |
| Variable: origin_PWM | Importance: 0.0 |
| Variable: origin_RAP | Importance: 0.0 |
| Variable: origin_RDU | Importance: 0.0 |
| Variable: origin_RFD | Importance: 0.0 |
| Variable: origin_RIC | Importance: 0.0 |
| Variable: origin_RNO | Importance: 0.0 |
| Variable: origin_ROA | Importance: 0.0 |
| Variable: origin_ROC | Importance: 0.0 |
| Variable: origin_SAN | Importance: 0.0 |
| Variable: origin_SAT | Importance: 0.0 |
| Variable: origin_SBN | Importance: 0.0 |

| | |
|----------------------|-----------------|
| Variable: origin_SCK | Importance: 0.0 |
| Variable: origin_SDF | Importance: 0.0 |
| Variable: origin_SEA | Importance: 0.0 |
| Variable: origin_SFO | Importance: 0.0 |
| Variable: origin_SGF | Importance: 0.0 |
| Variable: origin_SHV | Importance: 0.0 |
| Variable: origin_SLC | Importance: 0.0 |
| Variable: origin_SMX | Importance: 0.0 |
| Variable: origin_SPI | Importance: 0.0 |
| Variable: origin_STL | Importance: 0.0 |
| Variable: origin_SYR | Importance: 0.0 |
| Variable: origin_TLH | Importance: 0.0 |
| Variable: origin_TOL | Importance: 0.0 |
| Variable: origin_TPA | Importance: 0.0 |
| Variable: origin_TRI | Importance: 0.0 |
| Variable: origin_TUL | Importance: 0.0 |
| Variable: origin_TUS | Importance: 0.0 |
| Variable: origin_TYS | Importance: 0.0 |
| Variable: dest_ABE | Importance: 0.0 |
| Variable: dest_ABQ | Importance: 0.0 |
| Variable: dest_ACK | Importance: 0.0 |
| Variable: dest_ACY | Importance: 0.0 |
| Variable: dest_AGS | Importance: 0.0 |
| Variable: dest_ALB | Importance: 0.0 |
| Variable: dest_AMA | Importance: 0.0 |
| Variable: dest_ANC | Importance: 0.0 |
| Variable: dest_ATW | Importance: 0.0 |
| Variable: dest_AUS | Importance: 0.0 |
| Variable: dest_AVL | Importance: 0.0 |
| Variable: dest_AVP | Importance: 0.0 |
| Variable: dest_AZA | Importance: 0.0 |
| Variable: dest_BDL | Importance: 0.0 |
| Variable: dest_BGR | Importance: 0.0 |
| Variable: dest_BHM | Importance: 0.0 |
| Variable: dest_BIL | Importance: 0.0 |
| Variable: dest_BIS | Importance: 0.0 |
| Variable: dest_BLI | Importance: 0.0 |
| Variable: dest_BLV | Importance: 0.0 |
| Variable: dest_BMI | Importance: 0.0 |
| Variable: dest_BNA | Importance: 0.0 |
| Variable: dest_BOI | Importance: 0.0 |
| Variable: dest_BQN | Importance: 0.0 |
| Variable: dest_BTR | Importance: 0.0 |
| Variable: dest_BTV | Importance: 0.0 |
| Variable: dest_BUF | Importance: 0.0 |
| Variable: dest_BUR | Importance: 0.0 |
| Variable: dest_BWI | Importance: 0.0 |
| Variable: dest_BZN | Importance: 0.0 |
| Variable: dest_CAE | Importance: 0.0 |
| Variable: dest_CAK | Importance: 0.0 |
| Variable: dest_CHA | Importance: 0.0 |
| Variable: dest_CHO | Importance: 0.0 |
| Variable: dest_CHS | Importance: 0.0 |
| Variable: dest_CID | Importance: 0.0 |
| Variable: dest_CLE | Importance: 0.0 |
| Variable: dest_CLT | Importance: 0.0 |
| Variable: dest_CMH | Importance: 0.0 |
| Variable: dest_COS | Importance: 0.0 |
| Variable: dest_CRW | Importance: 0.0 |
| Variable: dest_CVG | Importance: 0.0 |
| Variable: dest_DAB | Importance: 0.0 |
| Variable: dest_DAL | Importance: 0.0 |
| Variable: dest_DAY | Importance: 0.0 |
| Variable: dest_DCA | Importance: 0.0 |
| Variable: dest_DLH | Importance: 0.0 |
| Variable: dest_DSM | Importance: 0.0 |
| Variable: dest_DTW | Importance: 0.0 |
| Variable: dest_ECP | Importance: 0.0 |
| Variable: dest_EGE | Importance: 0.0 |
| Variable: dest_ELP | Importance: 0.0 |
| Variable: dest_EUG | Importance: 0.0 |
| Variable: dest_EVV | Importance: 0.0 |
| Variable: dest_EWR | Importance: 0.0 |
| Variable: dest_EYW | Importance: 0.0 |
| Variable: dest_FAI | Importance: 0.0 |
| Variable: dest_FAR | Importance: 0.0 |
| Variable: dest_FAT | Importance: 0.0 |
| Variable: dest_FAY | Importance: 0.0 |
| Variable: dest_FCA | Importance: 0.0 |
| Variable: dest_FNT | Importance: 0.0 |
| Variable: dest_FSD | Importance: 0.0 |
| Variable: dest_GEG | Importance: 0.0 |
| Variable: dest_GIT | Importance: 0.0 |

```

variable. dest_GJI      importance. 0.0
Variable: dest_GNV      Importance: 0.0
Variable: dest_GPT      Importance: 0.0
Variable: dest_GRB      Importance: 0.0
Variable: dest_GRI      Importance: 0.0
Variable: dest_GRR      Importance: 0.0
Variable: dest_GSO      Importance: 0.0
Variable: dest_GSP      Importance: 0.0
Variable: dest_GTF      Importance: 0.0
Variable: dest_GUC      Importance: 0.0
Variable: dest_HDN      Importance: 0.0
Variable: dest_HNL      Importance: 0.0
Variable: dest_HOU      Importance: 0.0
Variable: dest_HPNI     Importance: 0.0
Variable: dest_HRL      Importance: 0.0
Variable: dest_HSV      Importance: 0.0
Variable: dest_HYA      Importance: 0.0
Variable: dest_IAD      Importance: 0.0
Variable: dest_IAG      Importance: 0.0
Variable: dest_IAH      Importance: 0.0
Variable: dest_ICT      Importance: 0.0
Variable: dest_IDA      Importance: 0.0
Variable: dest_ILG      Importance: 0.0
Variable: dest_ILM      Importance: 0.0
Variable: dest_IND      Importance: 0.0
Variable: dest_ISP      Importance: 0.0
Variable: dest_ITO      Importance: 0.0
Variable: dest_JAC      Importance: 0.0
Variable: dest_JAN      Importance: 0.0
Variable: dest_JAX      Importance: 0.0
Variable: dest_JNU      Importance: 0.0
Variable: dest_KOA      Importance: 0.0
Variable: dest_KTN      Importance: 0.0
Variable: dest_LAS      Importance: 0.0
Variable: dest_LAX      Importance: 0.0
Variable: dest_LBB      Importance: 0.0
Variable: dest_LBE      Importance: 0.0
Variable: dest_LCK      Importance: 0.0
Variable: dest_LEX      Importance: 0.0
Variable: dest_LFT      Importance: 0.0
Variable: dest_LGA      Importance: 0.0
Variable: dest_LGB      Importance: 0.0
Variable: dest_LIH      Importance: 0.0
Variable: dest_LIT      Importance: 0.0
Variable: dest_LRD      Importance: 0.0
Variable: dest_MAF      Importance: 0.0
Variable: dest_MCI      Importance: 0.0
Variable: dest_MDT      Importance: 0.0
Variable: dest_MDW      Importance: 0.0
Variable: dest_MEM      Importance: 0.0
Variable: dest_MFE      Importance: 0.0
Variable: dest_MFR      Importance: 0.0
Variable: dest_MHT      Importance: 0.0
Variable: dest_MIA      Importance: 0.0
Variable: dest_MKE      Importance: 0.0
Variable: dest_MLB      Importance: 0.0
Variable: dest_MLI      Importance: 0.0
Variable: dest_MOB      Importance: 0.0
Variable: dest_MOT      Importance: 0.0
Variable: dest_MRY      Importance: 0.0
Variable: dest_MSN      Importance: 0.0
Variable: dest_MSO      Importance: 0.0
Variable: dest_MSP      Importance: 0.0
Variable: dest_MSX      Importance: 0.0
Variable: dest_MTJ      Importance: 0.0
Variable: dest_MVY      Importance: 0.0
Variable: dest_MYR      Importance: 0.0
Variable: dest_OAJ      Importance: 0.0
Variable: dest_OAK      Importance: 0.0
Variable: dest_OGG      Importance: 0.0
Variable: dest_OKC      Importance: 0.0
Variable: dest_OMA      Importance: 0.0
Variable: dest_ONT      Importance: 0.0
Variable: dest_ORF      Importance: 0.0
Variable: dest_ORH      Importance: 0.0
Variable: dest_PBG      Importance: 0.0
Variable: dest_PBI      Importance: 0.0
Variable: dest_PDX      Importance: 0.0
Variable: dest_PGD      Importance: 0.0
Variable: dest_PHF      Importance: 0.0
Variable: dest_PHL      Importance: 0.0
Variable: dest_PHX      Importance: 0.0
Variable: dest_PIA      Importance: 0.0

```

```

Variable: dest_PIE      Importance: 0.0
Variable: dest_PIT      Importance: 0.0
Variable: dest_PNS      Importance: 0.0
Variable: dest_PSC      Importance: 0.0
Variable: dest_PSE      Importance: 0.0
Variable: dest_PSP      Importance: 0.0
Variable: dest_PVD      Importance: 0.0
Variable: dest_PVU      Importance: 0.0
Variable: dest_PWM      Importance: 0.0
Variable: dest_RAP      Importance: 0.0
Variable: dest_RDU      Importance: 0.0
Variable: dest_RFD      Importance: 0.0
Variable: dest_RIC      Importance: 0.0
Variable: dest_RNO      Importance: 0.0
Variable: dest_ROA      Importance: 0.0
Variable: dest_ROC      Importance: 0.0
Variable: dest_RSW      Importance: 0.0
Variable: dest_SAN      Importance: 0.0
Variable: dest_SAT      Importance: 0.0
Variable: dest_SAV      Importance: 0.0
Variable: dest_SBA      Importance: 0.0
Variable: dest_SCK      Importance: 0.0
Variable: dest_SDF      Importance: 0.0
Variable: dest_SEA      Importance: 0.0
Variable: dest_SFB      Importance: 0.0
Variable: dest_SFO      Importance: 0.0
Variable: dest_SGF      Importance: 0.0
Variable: dest_SHV      Importance: 0.0
Variable: dest_SIT      Importance: 0.0
Variable: dest_SJC      Importance: 0.0
Variable: dest_SJU      Importance: 0.0
Variable: dest_SLC      Importance: 0.0
Variable: dest_SMF      Importance: 0.0
Variable: dest_SNA      Importance: 0.0
Variable: dest_SRQ      Importance: 0.0
Variable: dest_STL      Importance: 0.0
Variable: dest_STT      Importance: 0.0
Variable: dest_STX      Importance: 0.0
Variable: dest_SWF      Importance: 0.0
Variable: dest_SYR      Importance: 0.0
Variable: dest_TLH      Importance: 0.0
Variable: dest_TPA      Importance: 0.0
Variable: dest_TRI      Importance: 0.0
Variable: dest_TTN      Importance: 0.0
Variable: dest_TUL      Importance: 0.0
Variable: dest_TUS      Importance: 0.0
Variable: dest_TVC      Importance: 0.0
Variable: dest_TYS      Importance: 0.0
Variable: dest_USA      Importance: 0.0
Variable: dest_UST      Importance: 0.0
Variable: dest_VPS      Importance: 0.0
Variable: dest_XNA      Importance: 0.0

```

```

rf_dep = pd.DataFrame(feature_importances).head(10)
rf_dep.columns = ["Variables", "Importance"]
rf_dep

```

| | Variables | Importance |
|---|--------------------|------------|
| 0 | distance | 0.07 |
| 1 | temp | 0.07 |
| 2 | dew_point_temp | 0.07 |
| 3 | wind_speed | 0.07 |
| 4 | sea_level_pressure | 0.06 |
| 5 | station_pressure | 0.06 |
| 6 | visibilty | 0.04 |
| 7 | prcp | 0.03 |
| 8 | rain_drizzle | 0.01 |
| 9 | thunder | 0.01 |

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(20,5))
```

```
sns.set(style="whitegrid")
```

```
ax = sns.barplot(x="Variables", y="Importance", data=rf_dep).set_title("Importance Variables of Departure Delay", fontsize = 20)
```

