

```
pip install pandas
```

```
Requirement already satisfied: pandas in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (2.2.3)  
Requirement already satisfied: numpy>=1.26.0 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from pandas) (2.2.0)  
Requirement already satisfied: python-dateutil>=2.8.2 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from pandas) (2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from pandas) (2024.2)  
Requirement already satisfied: tzdata>=2022.7 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from pandas) (2024.2)  
Requirement already satisfied: six>=1.5 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
```

```
[notice] A new release of pip is available: 24.2 -> 24.3.1
```

```
[notice] To update, run:
```

```
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/bin/python -m pip  
install --upgrade pip
```

```
Note: you may need to restart the kernel to use updated packages.
```

```
import pandas as pd
```

```
import numpy as np
```

```
pip install matplotlib
```

```
Requirement already satisfied: matplotlib in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (3.10.0)  
Requirement already satisfied: contourpy>=1.0.1 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from matplotlib) (1.3.1)  
Requirement already satisfied: cycler>=0.10 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from matplotlib) (0.12.1)  
Requirement already satisfied: fonttools>=4.22.0 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from matplotlib) (4.55.3)  
Requirement already satisfied: kiwisolver>=1.3.1 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-  
packages (from matplotlib) (1.4.7)  
Requirement already satisfied: numpy>=1.23 in  
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
```

```
packages (from matplotlib) (2.2.0)
Requirement already satisfied: packaging>=20.0 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: six>=1.5 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from python-dateutil>=2.7->matplotlib) (1.16.0)
```

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run:

```
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/bin/python -m pip
install --upgrade pip
```

Note: you may need to restart the kernel to use updated packages.

```
pip install seaborn
```

```
Requirement already satisfied: seaborn in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (0.13.2)
Requirement already satisfied: numpy!=1.24.0,>=1.20 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from seaborn) (2.2.0)
Requirement already satisfied: pandas>=1.2 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from seaborn) (2.2.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from seaborn) (3.10.0)
Requirement already satisfied: contourpy>=1.0.1 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.3.1)
Requirement already satisfied: cycler>=0.10 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.55.3)
Requirement already satisfied: kiwisolver>=1.3.1 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.7)
Requirement already satisfied: packaging>=20.0 in
```

```
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.2)
Requirement already satisfied: pillow>=8 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from pandas>=1.2->seaborn) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from pandas>=1.2->seaborn) (2024.2)
Requirement already satisfied: six>=1.5 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn)
(1.16.0)
```

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run:

```
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/bin/python -m pip
install --upgrade pip
```

Note: you may need to restart the kernel to use updated packages.

```
pip install -U scikit-learn
```

```
Requirement already satisfied: scikit-learn in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (1.6.0)
Requirement already satisfied: numpy>=1.19.5 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from scikit-learn) (2.2.0)
Requirement already satisfied: scipy>=1.6.0 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from scikit-learn) (1.14.1)
Requirement already satisfied: joblib>=1.2.0 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/lib/python3.12/site-
packages (from scikit-learn) (3.5.0)
```

[notice] A new release of pip is available: 24.2 -> 24.3.1

[notice] To update, run:

```
/opt/homebrew/Cellar/jupyterlab/4.3.1_1/libexec/bin/python -m pip
```

```
install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import LabelEncoder
```

```
# Load the datasets
```

```
train_data = pd.read_csv("./train.csv")
```

```
test_data = pd.read_csv("./test.csv")
```

```
pincodes_data = pd.read_csv("./pincodes.csv", low_memory=False)
```

```
# Explore data
```

```
print(train_data.head())
```

	id	order_delivered_date	courier_partner_id	order_shipped_date \
0	349197554	2022-06-05	9	2022-06-02
1	349197557	2022-06-04	9	2022-06-02
2	349197560	2022-06-07	9	2022-06-02
3	349197565	2022-06-05	9	2022-06-02
4	349197568	2022-06-07	9	2022-06-02

	account_type_id	drop_pin_code	pickup_pin_code	quantity
0	2471	141007	421311	1
1	2471	411047	421311	1
2	2471	145029	421311	1
3	2471	751015	562123	1
4	2471	145029	421311	1

	order_delivery_sla
0	3
1	2
2	5
3	3
4	5

```
print(test_data.head())
```

	id	courier_partner_id	order_shipped_date	account_type_id \
0	428365149	3	2022-09-02	2520
1	428365152	3	2022-09-02	2520
2	428365154	3	2022-09-02	2520
3	428365156	3	2022-09-02	2520
4	428365161	3	2022-09-02	2520

	drop_pin_code	pickup_pin_code	quantity	account_mode
0	243001	122506	1	Air
1	209502	122506	1	Air
2	147001	122506	1	Air
3	245101	122506	1	Air
4	305901	122506	1	Air

```
print(pincodes_data.head())
```

	CircleName	RegionName	DivisionName \
0	Andhra Pradesh Circle	Kurnool Region	Hindupur Division
1	Andhra Pradesh Circle	Kurnool Region	Hindupur Division
2	Andhra Pradesh Circle	Kurnool Region	Hindupur Division
3	Andhra Pradesh Circle	Kurnool Region	Hindupur Division
4	Andhra Pradesh Circle	Kurnool Region	Hindupur Division

	OfficeName	Pincode	OfficeType	Delivery	District \
0	Peddakotla B.0	515631	B0	Delivery	ANANTAPUR
1	Pinnadhari B.0	515631	B0	Delivery	ANANTAPUR
2	Yerraguntapalle B.0	515631	B0	Delivery	ANANTAPUR
3	Obulareddipalli B.0	515581	B0	Delivery	ANANTAPUR
4	Odulapalli B.0	515581	B0	Delivery	ANANTAPUR

	StateName	Latitude	Longitude
0	ANDHRA PRADESH	14.5689	77.85624
1	ANDHRA PRADESH	14.5281	77.857014
2	ANDHRA PRADESH	14.561111	77.85715
3	ANDHRA PRADESH	14.2488	78.2588
4	ANDHRA PRADESH	14.24555	78.2477

```
# Data Preprocessing
```

```
train_data['order_shipped_date'] =  
pd.to_datetime(train_data['order_shipped_date'])
```

```
train_data['order_delivered_date'] =  
pd.to_datetime(train_data['order_delivered_date'])
```

```

# Feature engineering: calculate actual SLA in training set
train_data['actual_sla'] = (train_data['order_delivered_date'] -
train_data['order_shipped_date']).dt.days

from sklearn.preprocessing import LabelEncoder

# Encode categorical variables
categorical_cols = ['courier_partner_id', 'account_type_id',
'account_mode']
label_encoders = {}

for col in categorical_cols:
    le = LabelEncoder()

    # Fit the LabelEncoder on train data
    train_data[col] = le.fit_transform(train_data[col])
    label_encoders[col] = le

    # Transform test data, handling unseen labels
    test_data[col] = test_data[col].apply(lambda x: le.transform([x])
[0] if x in le.classes_ else -1)

# Feature selection
features = [
    'order_shipped_date', 'quantity', 'courier_partner_id',
    'account_type_id', 'drop_pin_code', 'pickup_pin_code',
    'account_mode'
]
target = 'actual_sla'

# Convert dates into numerical values for the model
train_data['order_shipped_date'] =
train_data['order_shipped_date'].astype(int) / 10**9
test_data['order_shipped_date'] =
pd.to_datetime(test_data['order_shipped_date']).astype(int) / 10**9

# Split data for training and validation
X = train_data[features]
y = train_data[target]
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2,
random_state=42)

# Model training
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

RandomForestRegressor(random_state=42)

# Model validation
y_pred = rf_model.predict(X_val)

```

```
rmse = np.sqrt(mean_squared_error(y_val, y_pred))
print(f"Validation RMSE: {rmse}")
```

Validation RMSE: 0.7601252529590619

```
from sklearn.metrics import mean_absolute_error
mae = mean_absolute_error(y_val, y_pred)
print(f'Validation MAE: {mae}')
```

Validation MAE: 0.36453799423135036

```
from sklearn.metrics import r2_score
r2 = r2_score(y_val, y_pred)
print(f'Validation R²: {r2}')
```

Validation R²: 0.8074559915431081

Feature importance

```
feature_importances = pd.DataFrame({
    'Feature': features,
    'Importance': rf_model.feature_importances_
}).sort_values(by='Importance', ascending=False)
print(feature_importances)
```

	Feature	Importance
2	courier_partner_id	0.373501
4	drop_pin_code	0.366301
0	order_shipped_date	0.146914
5	pickup_pin_code	0.087783
6	account_mode	0.018469
3	account_type_id	0.007031
1	quantity	0.000000

Test prediction

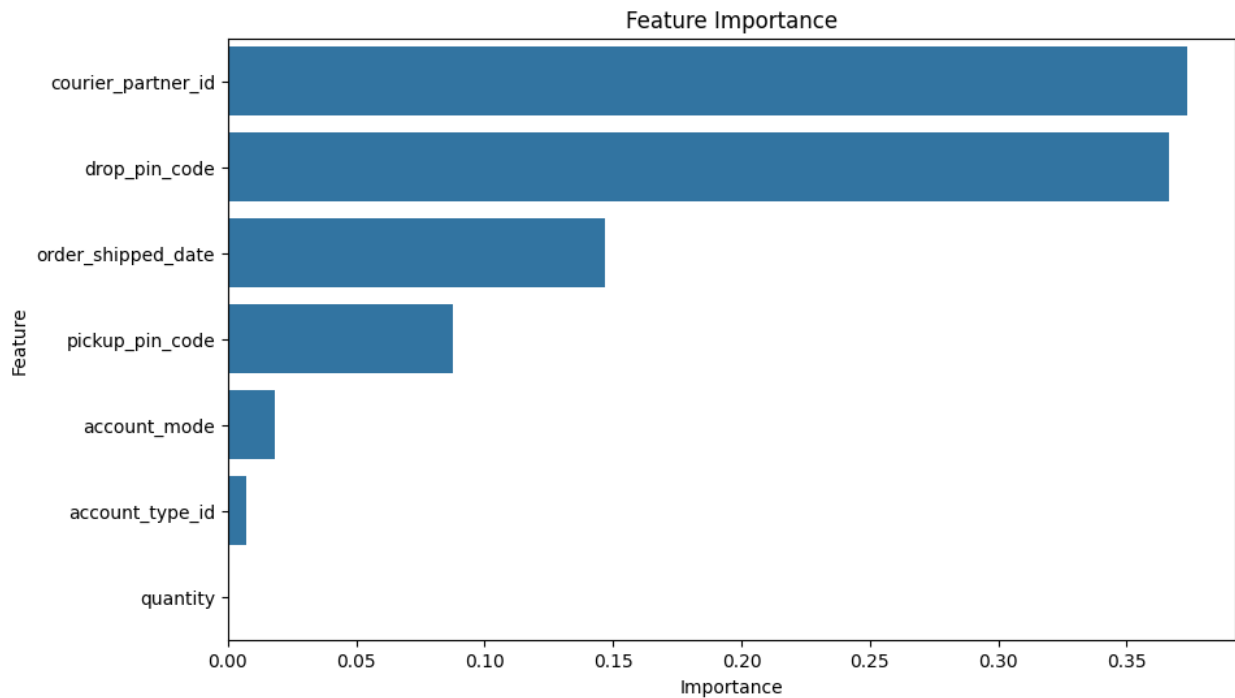
```
test_data['predicted_exact_sla'] =
rf_model.predict(test_data[features])
```

Save submission

```
submission = test_data[['id', 'predicted_exact_sla']]
submission.to_csv("submission.csv", index=False)
```

Visualization of feature importance

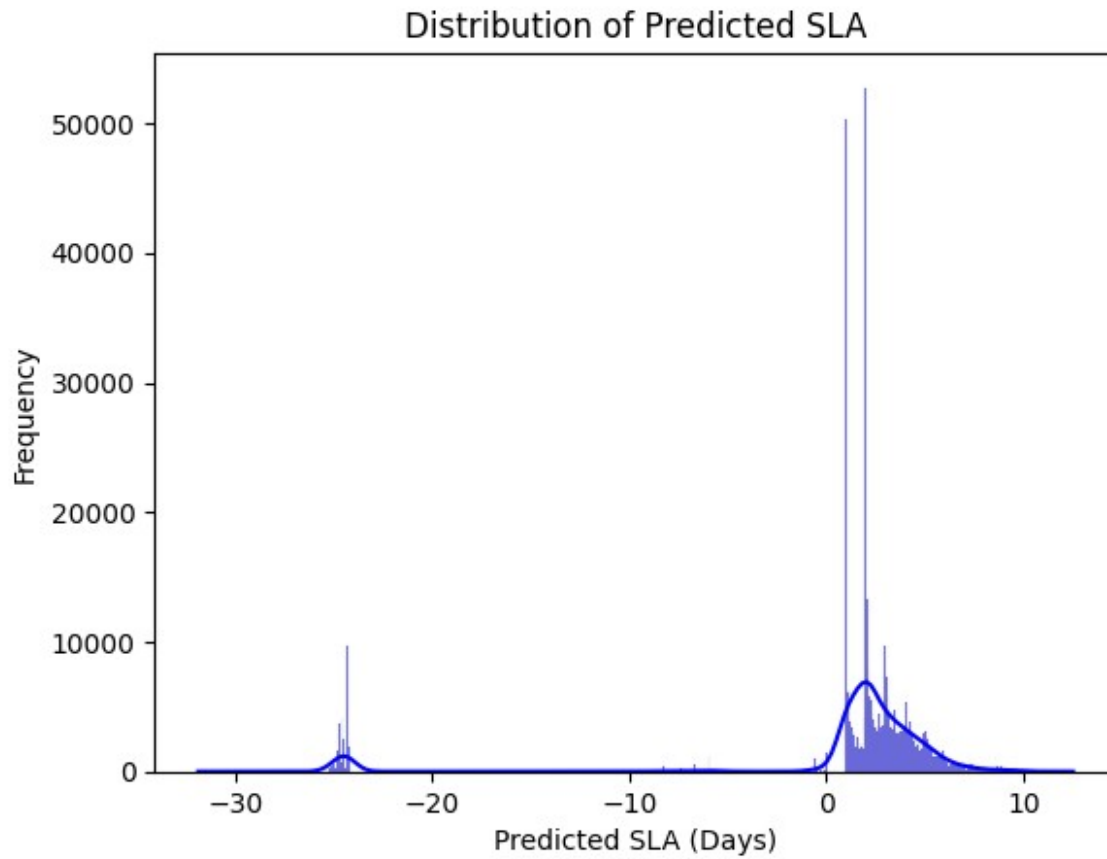
```
plt.figure(figsize=(10, 6))
sns.barplot(x='Importance', y='Feature', data=feature_importances)
plt.title('Feature Importance')
plt.show()
```



```
import seaborn as sns

# Load submission.csv
predictions = pd.read_csv('submission.csv')

# Plot SLA distribution
sns.histplot(predictions['predicted_exact_sla'], kde=True,
color='blue')
plt.title('Distribution of Predicted SLA')
plt.xlabel('Predicted SLA (Days)')
plt.ylabel('Frequency')
plt.savefig('sla_distribution.png')
plt.show()
```

```
residuals = y_val - y_pred
sns.histplot(residuals, kde=True)
plt.title("Residual Distribution")
plt.xlabel("Residuals")
plt.ylabel("Frequency")
plt.show()
```

