Выполнил – Гудковский Артём

1.

import pandas as pd  
import matplotlib.pyplot as plt  
  
# Загрузка данных  
df = pd.read\_csv('flight\_delays.csv')  
  
# Преобразование столбца 'dep\_delayed\_15min' в бинарный формат  
df['dep\_delayed\_15min'] = df['dep\_delayed\_15min'].map({'Y': 1, 'N': 0})  
  
# Расчет доли задержек ко всем вылетам  
delay\_ratio = df['dep\_delayed\_15min'].mean()  
  
# Визуализация  
plt.bar(['Задержки', 'Без задержек'], [delay\_ratio, 1 - delay\_ratio])  
plt.title('Доля задержек ко всем вылетам')  
plt.show()

2.

import matplotlib.pyplot as plt  
import seaborn as sns  
  
# Преобразование столбца 'Distance' в числовой формат  
df['Distance'] = pd.to\_numeric(df['Distance'], errors='coerce')  
  
# Создаем новый столбец 'Distance\_Group' с группировкой по 250 милям  
df['Distance\_Group'] = ((df['Distance'] + 249) // 250 \* 250).astype(int)  
  
# Группировка по длине пути и подсчет частоты задержек  
delay\_counts = df.groupby(['Distance\_Group', 'dep\_delayed\_15min']).size().unstack().fillna(0)  
  
# Визуализация  
delay\_counts.plot(kind='bar', stacked=True, color=['blue', 'red'])  
plt.title('Групповая столбчатая диаграмма: Доли задержек по длине пути')  
plt.xlabel('Длина пути (в милях)')  
plt.ylabel('Количество рейсов')  
plt.show()

3.

# Группировка по направлению и подсчет частоты задержек  
top\_destinations = df.groupby('Dest')['dep\_delayed\_15min'].mean().sort\_values(ascending=False).head(5)  
  
# Визуализация  
top\_destinations.plot(kind='bar')  
plt.title('Топ 5 направлений с наибольшей долей задержек')  
plt.show()

4.

# Преобразование столбца 'Month' в числовой формат  
df['Month'] = df['Month'].str.extract('c-(\d+)').astype(int)  
  
# Группировка по месяцу и подсчет частоты задержек  
delays\_by\_month = df.groupby('Month')['dep\_delayed\_15min'].mean()  
  
# Визуализация  
delays\_by\_month.plot(kind='bar')  
plt.title('Доли задержек по месяцам')  
plt.show()

Результаты:  
1.



2.



3.



4.

