Data Analysis of Kazakhstan Road Accidents

Using Python Language liblaries

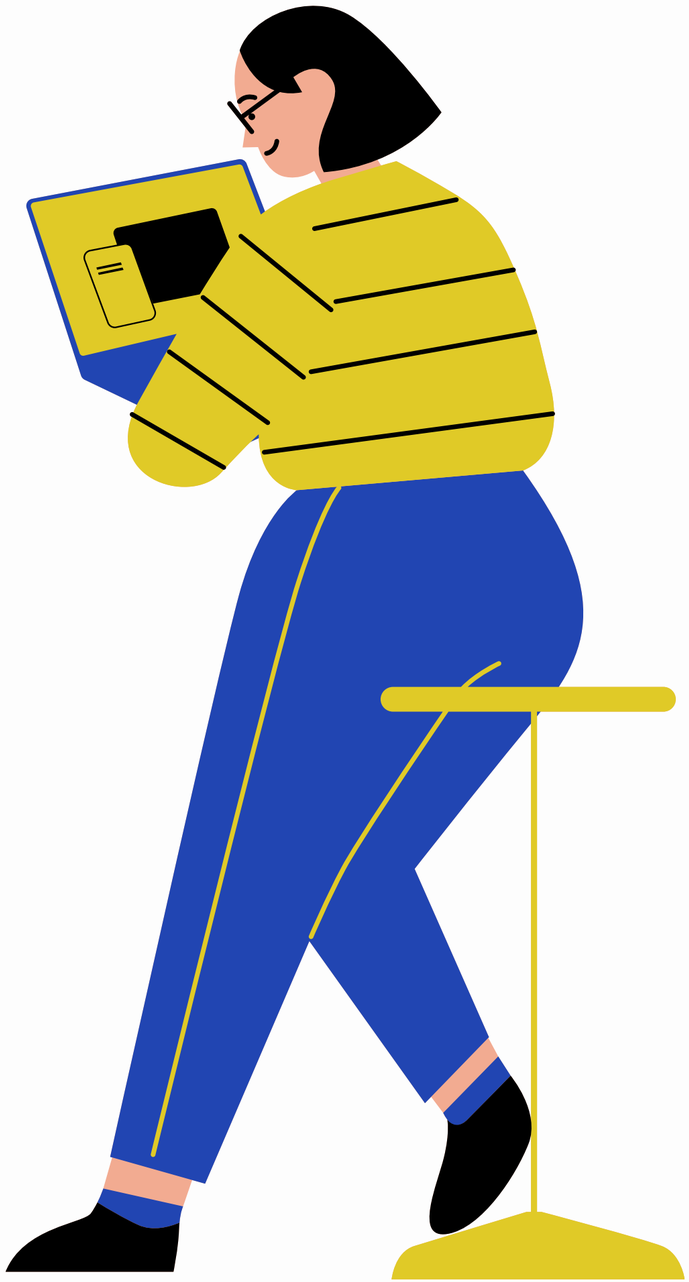
Garayev Emirdin

01 - Introduction

02 - Project structure

03 - Data Visualization

04 - Conclusion



Data

Analysis

Introduction

The choice fell on the topic of analyzing data on road traffic accidents (RTA) in Kazakhstan, since this is a key aspect of public safety and infrastructure development of our state. It is important to understand the causes, circumstances and trends of road accidents in order to take effective measures to prevent and reduce their impact on our society.

Studying traffic accident data: The path to safe roads and caring for everyone's lives.

Data

Analysis

Project structure

1-1. Filling "Privileges" column
1-2. Filling "Accident_region" column based on "City" & removing outliers
1-3. Filling "Loss_amount" column referring to the average losses

2-1. Analysis of accident frequency depending on age group
2-2. Analysis of accident frequency depending on driving experience
2-3. Analysis of accident frequency depending on gender group
2-4. Analysis of accident frequency depending on region
2-5. Analysis of accident frequency depending on steering wheel position
2-6. Analysis of accident frequency depending on citizenship (Without Kazakhstan)
2-7. Analysis of accident frequency depending on the vehicle (excluding cars)
2-8. Analysis of accident frequency depending on vehicle brands
2-9. Analysis of accident frequency depending on color
2-10. Analysis of accident frequency depending on benefits
2-11. Analysis of descriptive statistics

3-0. Final report

Preparing the table

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
excel_file_path = r'C:\Users\user\Desktop\kz_road_accident\kz_road_accident.xlsx'
df = pd.read_excel(excel_file_path)
df
```

1-1. Filling "Privileges" column

```
df['Privileges'].fillna('No privileges', inplace=True)
df[['Privileges']]
```

1-2. Filling "Accident_region" column based on "City"

```
df.loc[df['Accident_region'].isnull(), 'Accident_region'] = df['City']
df[['City', 'Accident_region']]
```

1-2. Removing outliers from the "Accident_region" column.

```
region_counts = df['Accident_region'].value_counts()
threshold = 20
df = df[df['Accident_region'].isin(region_counts[region_counts >= threshold].index)]
```

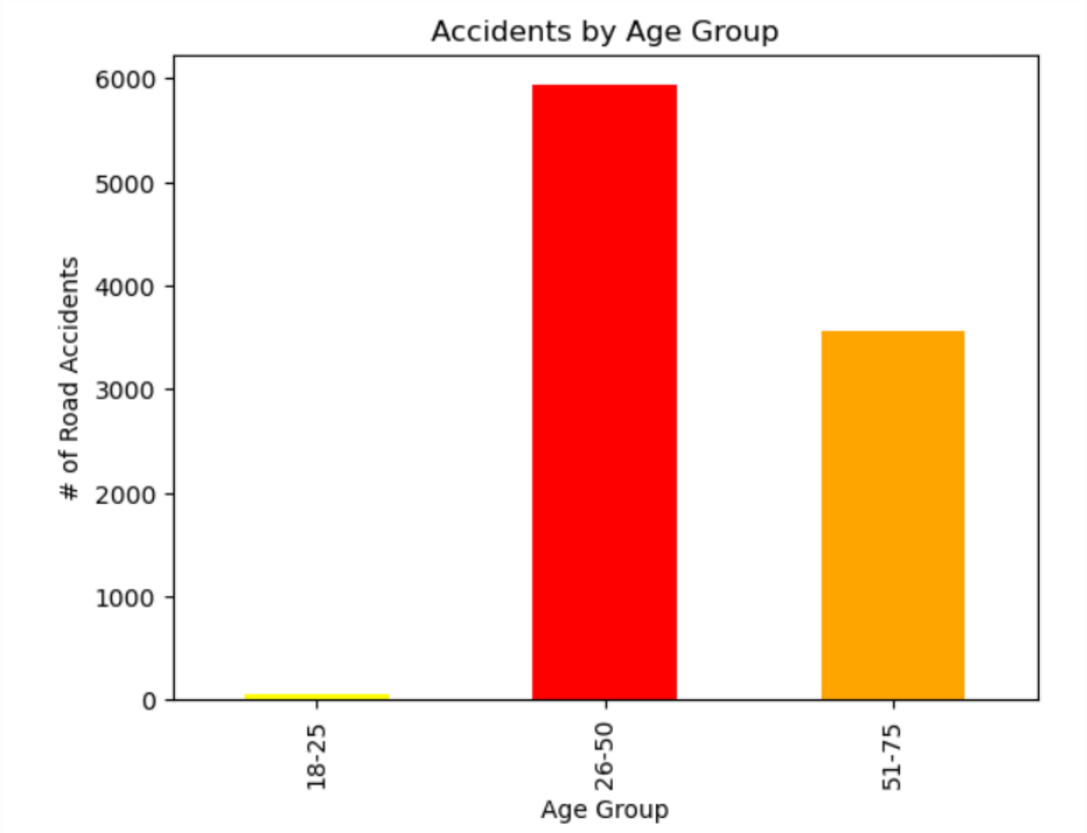
1-3. Filling "Loss_amount" column referring to the average losses

```
amount_of_loss = df['Loss_amount'].mean().round()
df.loc[df['Loss_amount'].isnull(), 'Loss_amount'] = amount_of_loss
df[['Loss_amount']]
```

Data analysis - Age group

```
bins = [18, 25, 50, 75]
labels = ['18-25', '26-50', '51-75']
df.loc[:, 'Age_group'] = pd.cut(df['Age'], bins=bins, labels=labels)
colors = ['yellow', 'red', 'orange']
age_group_counts = df['Age_group'].value_counts().sort_index()
age_group_counts.plot(kind='bar', xlabel='Age Group', ylabel='# of Road Accidents', title='Accidents by Age Group',
plt.show()
```

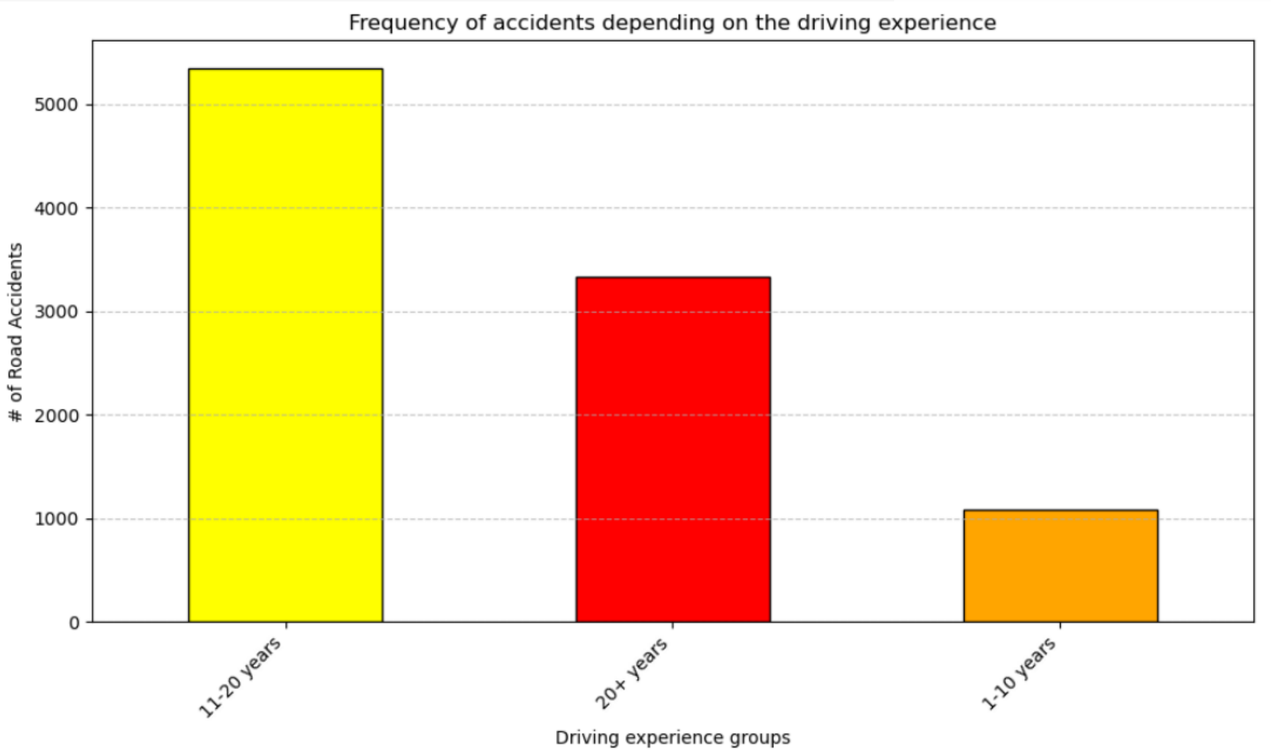
It is interesting that the 18-25 year old group, despite their activity on the roads, is involved in road accidents much less frequently, with only 56 cases. On the other hand, the age categories 26-50 and 51-75 years are noticeably predominant with 5929 and 3559 cases respectively. These data may indicate different levels of experience and risk awareness among different age groups on Kazakhstan's roads, but age does not equal driving experience.



Data analysis - Driving experience

```
bins = [0, 10, 20, float('inf')]
labels = ['1-10 years', '11-20 years', '20+ years']
colors = ['yellow', 'red', 'orange']

df.loc[:, 'Driving_experience_group'] = pd.cut(df['Driving_experience'], bins=bins, labels=labels, right=False)
accident_counts = df['Driving_experience_group'].value_counts()
plt.figure(figsize=(10, 6))
accident_counts.plot(kind='bar', color=colors, edgecolor='black')
plt.xlabel('Driving experience groups')
plt.ylabel('# of Road Accidents')
plt.title('Frequency of accidents depending on the driving experience')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

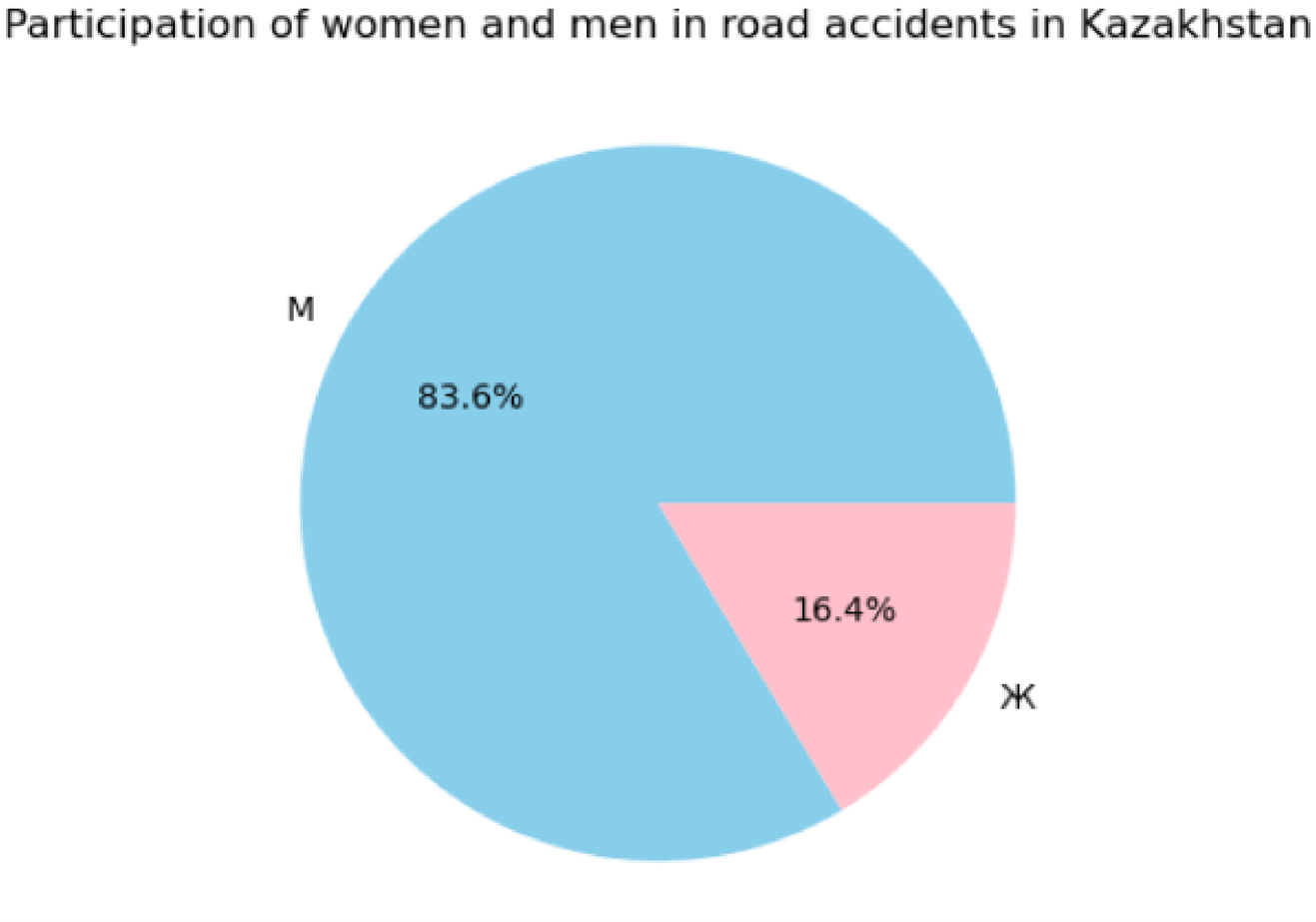
It is interesting to note that drivers with 11 to 20 years of driving experience account for the largest share of road accident statistics, with a total of 5,345 cases. While drivers with more than 20 years of experience are involved in road accidents less often, their number is 3,328 cases. However, drivers with 1 to 10 years of experience are also not left out, participating in 1,088 cases. These findings may indicate the importance of additional education and awareness of road safety, regardless of driving experience.



Data analysis - Gender group

```
gender_counts = df['Gender'].value_counts()
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%', colors=['skyblue', 'pink'])
plt.title('Participation of women and men in road accidents in Kazakhstan')
plt.show()
```

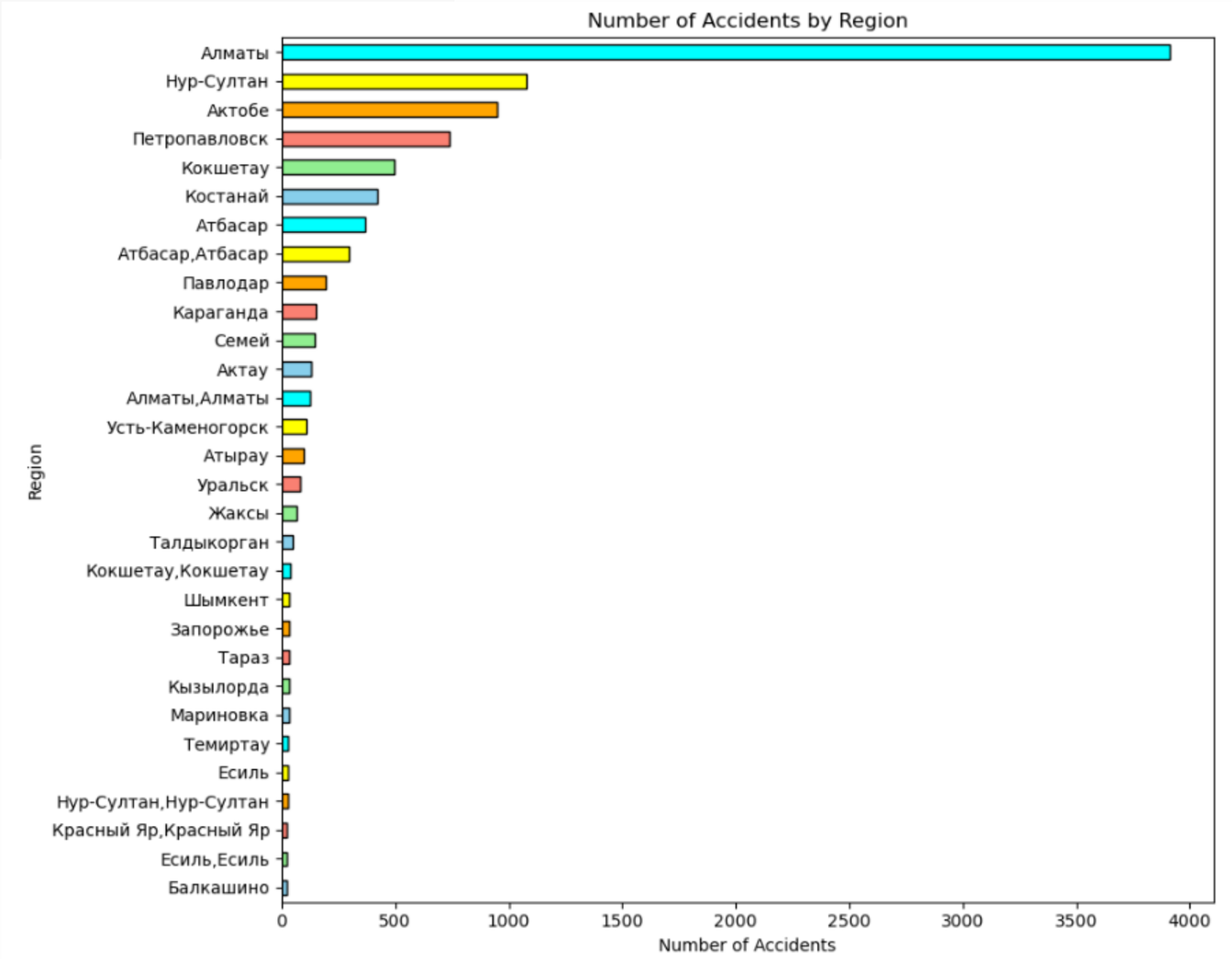
The statistics of road accidents are dominated by men, the total number of cases involving them is 8160. While women are involved in road accidents much less frequently, their number is only 1601 cases. These findings may indicate differences in driving styles and indicate that road safety aspects require further research and discussion in a gender-sensitive manner.



Data analysis - Regions

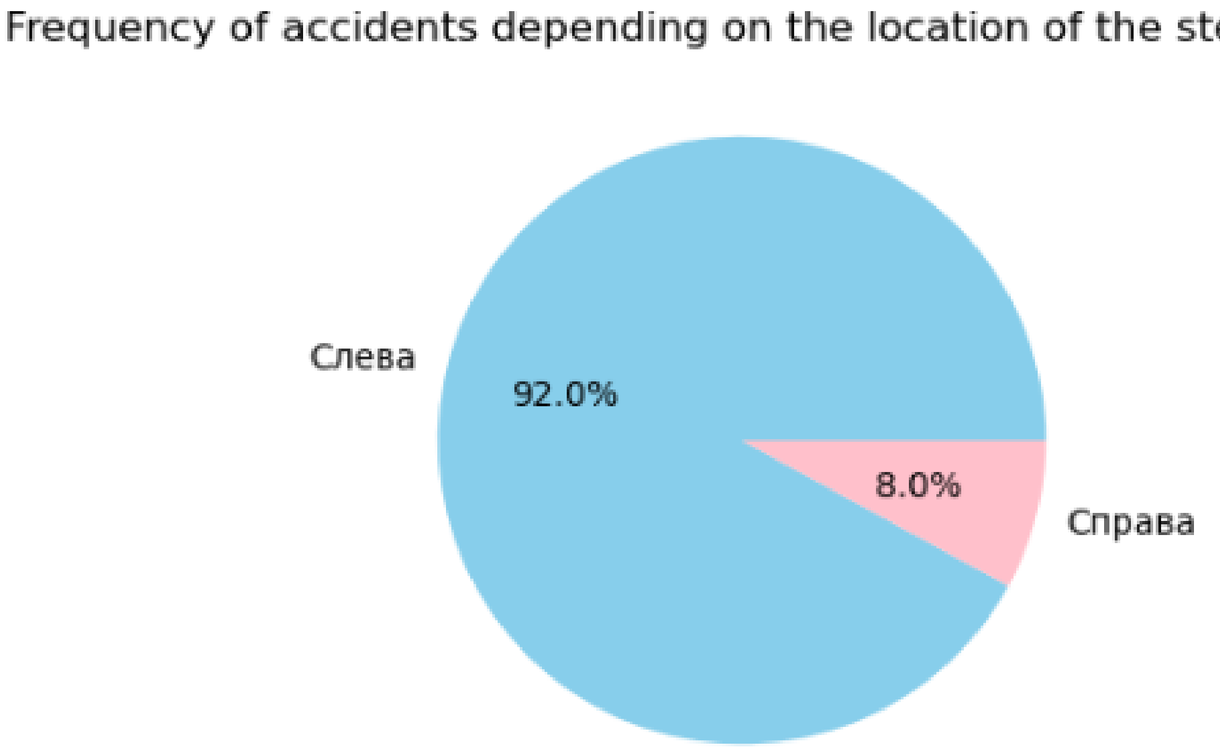
```
colors = ['skyblue', 'lightgreen', 'salmon', 'orange', 'yellow', 'cyan']
region_counts = df['Accident_region'].value_counts()
plt.figure(figsize=(10, 8))
region_counts.sort_values().plot(kind='barh', color=colors, edgecolor='black')
plt.xlabel('Number of Accidents')
plt.ylabel('Region')
plt.title('Number of Accidents by Region')
plt.tight_layout()
plt.show()
```

In the analysis of the regions of road accidents, the largest number of incidents was noted in Almaty, amounting to 3910 cases. It is followed by Astana with 1,080 cases and Aktobe with 947 cases. These data indicate the need for a more in-depth study of the causes and factors affecting road safety in these regions. Further analyzes can help identify the main problem areas and develop appropriate measures to improve them and reduce the number of accidents.



Data analysis - Wheel

```
accident_counts = df['Steering_wheel_location'].value_counts()
plt.figure(figsize=(4, 4))
plt.pie(accident_counts, labels=accident_counts.index, autopct='%1.1f%%', colors=['skyblue', 'pink'])
plt.title('Frequency of accidents depending on the location of the steering wheel')
plt.show()
```

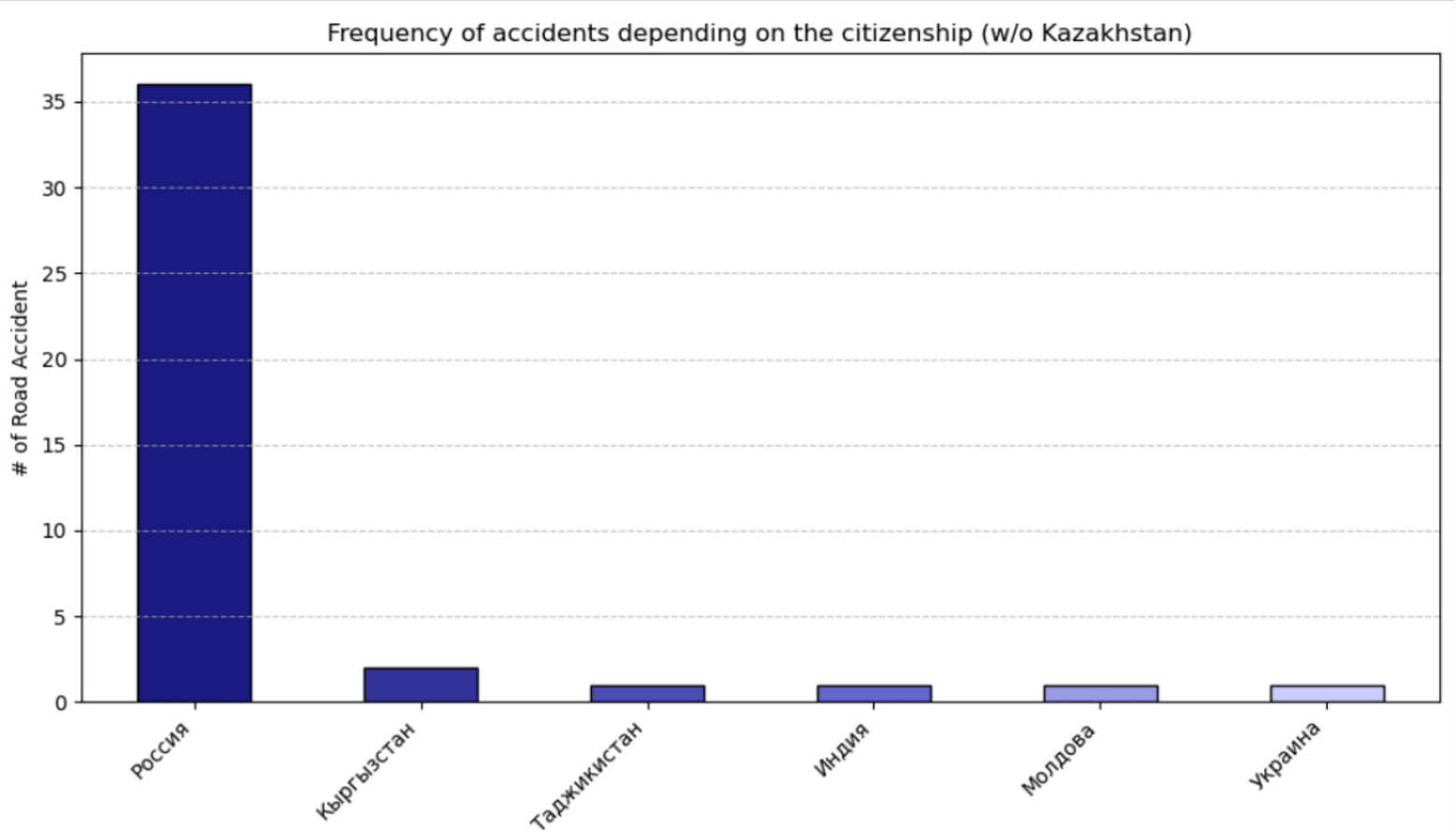


When studying data on the location of the steering wheel in road accidents, it is important to note that in Kazakhstan, traffic is on the right hand side. Despite this, left-hand drive vehicles still dominate accident statistics. Therefore, these data may indicate additional aspects of road safety that need to be considered and analyzed to develop appropriate measures to improve road safety in Kazakhstan.

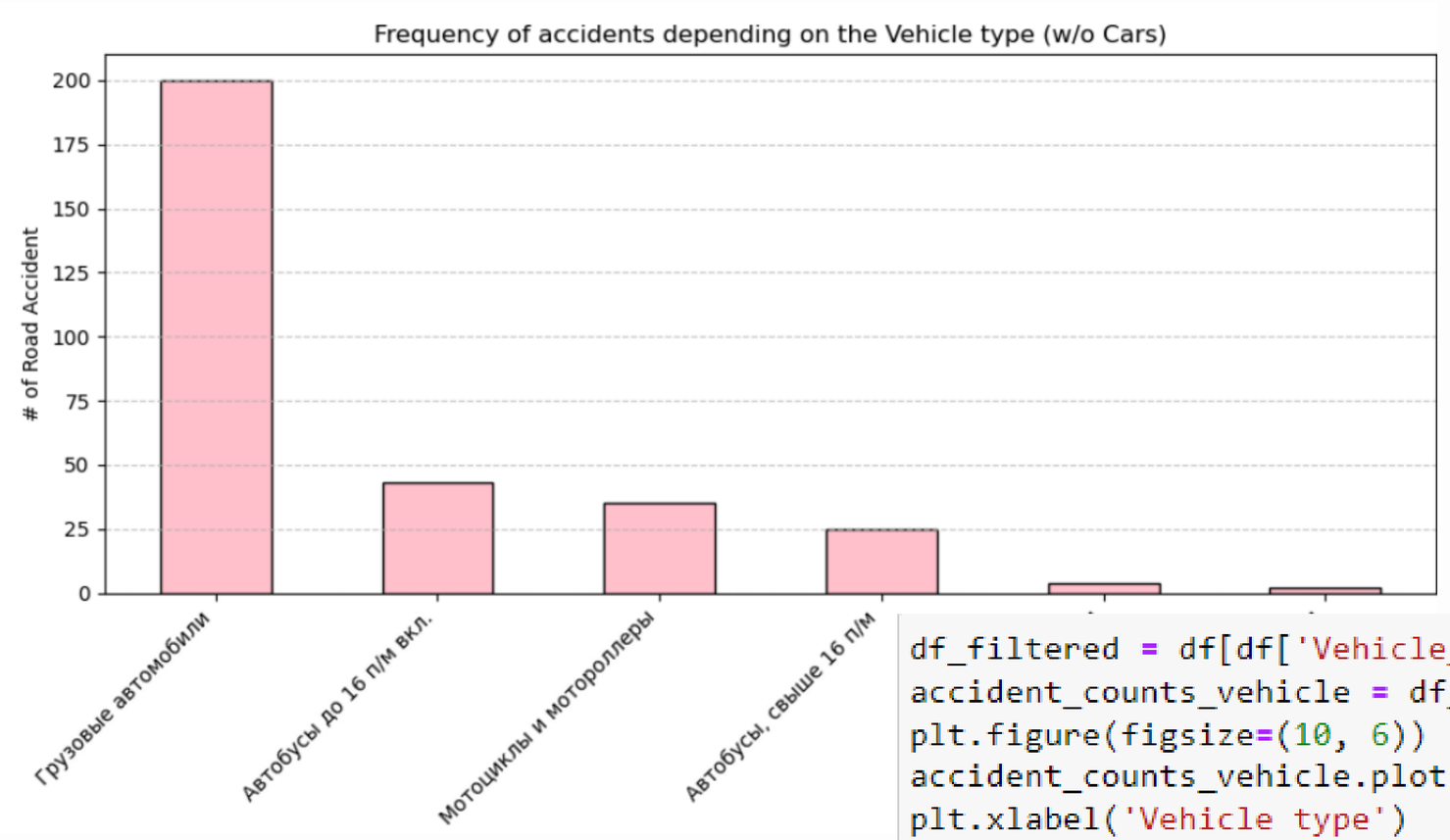
Data analysis - Citizenship

```
df_filtered = df[df['Citizenship'] != 'Казахстан']
accident_counts = df_filtered['Citizenship'].value_counts()
plt.figure(figsize=(10, 6))
accident_counts.plot(kind='bar', color=colors, edgecolor='black')
plt.xlabel('Citizenship')
plt.ylabel('# of Road Accident')
plt.title('Frequency of accidents depending on the citizenship (w/o Kazakhstan)')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

In the data presented on citizenship, citizens of Russia predominate, with the number of cases equal to 36. However, it is worth considering that this list does not include citizenship of Kazakhstan, which is an outlier.



Data analysis - Vehicle



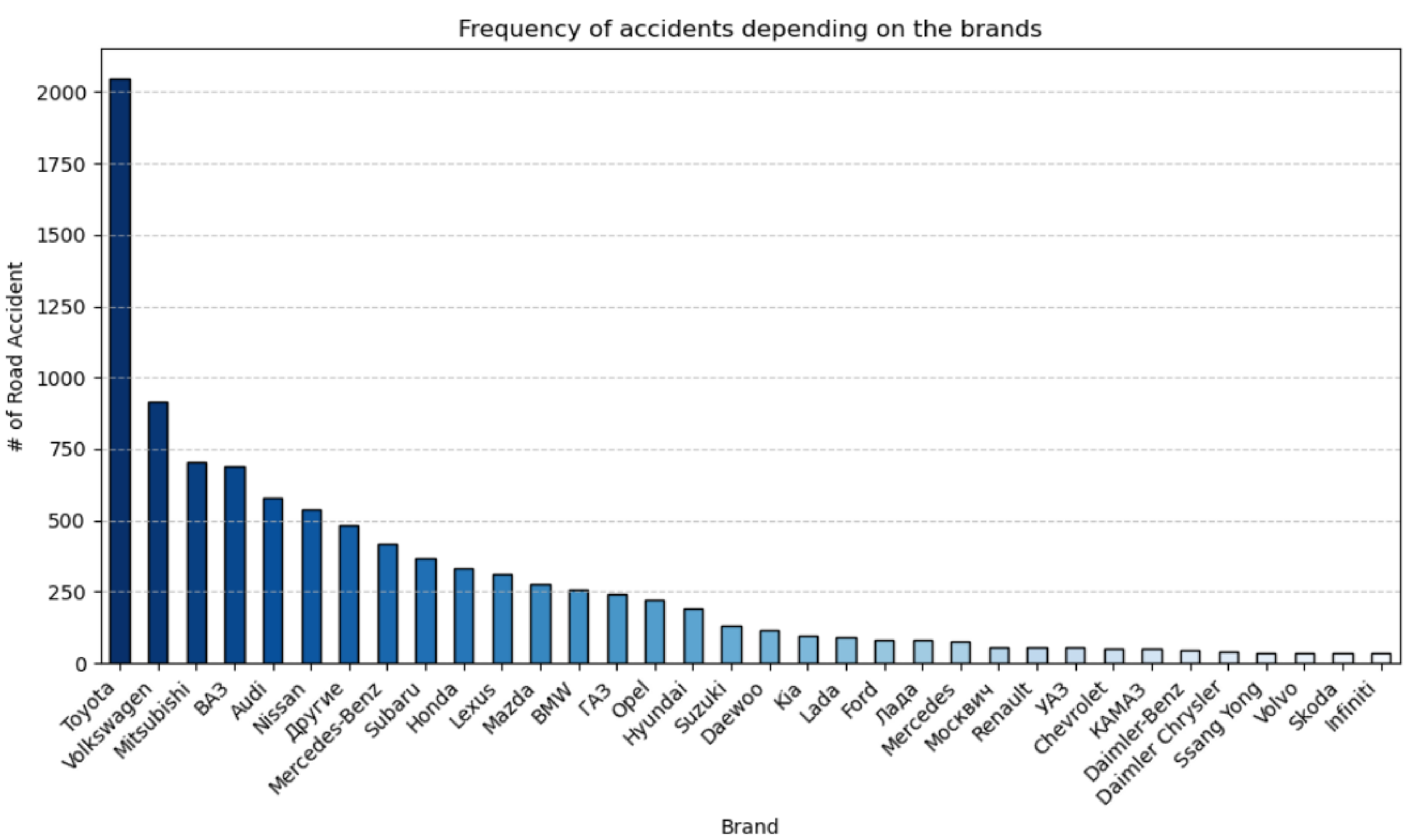
The vehicle type data reported is dominated by trucks with 200 cases, followed by buses with up to 16 passenger seats and motorcycles with 43 and 35 cases respectively. It is important to note that passenger cars are not mentioned in this list.

```
df_filtered = df[df['Vehicle_type '] != 'Легковые автомобили']
accident_counts_vehicle = df_filtered['Vehicle_type '].value_counts()
plt.figure(figsize=(10, 6))
accident_counts_vehicle.plot(kind='bar', color='pink', edgecolor='black')
plt.xlabel('Vehicle type')
plt.ylabel('# of Road Accident')
plt.title('Frequency of accidents depending on the Vehicle type (w/o Cars)')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

Data analysis - Brands

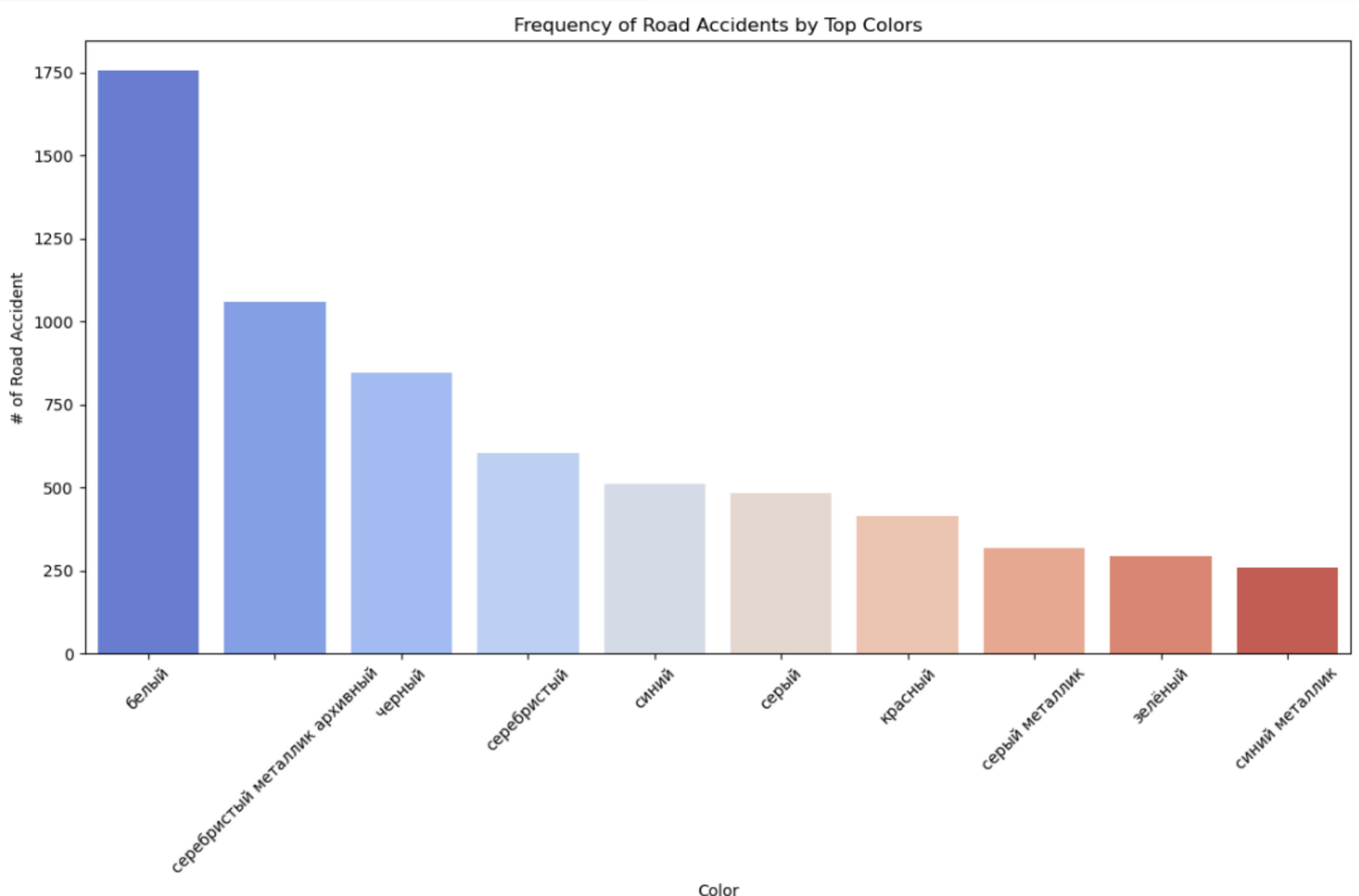
```
accident_counts_brand = df['Brand'].value_counts()
rare_brands = accident_counts_brand[accident_counts_brand < 30].index.tolist()
df.loc[df['Brand'].isin(rare_brands), 'Brand'] = 'Другие'
accident_counts_brand_updated = df['Brand'].value_counts()
colors = plt.cm.Blues(np.linspace(1, 0, len(accident_counts_brand_updated)))

plt.figure(figsize=(10, 6))
accident_counts_brand_updated.plot(kind='bar', color=colors, edgecolor='black')
plt.xlabel('Brand')
plt.ylabel('# of Road Accident')
plt.title('Frequency of accidents depending on the brands')
plt.xticks(rotation=45, ha='right')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()
```

The Toyota brand demonstrates significant superiority, with more than 2,000 cases. In second place are Volkswagen and Mitsubishi, each with more than 800 cases. These data may indicate the prevalence and availability of cars of these brands on the roads of Kazakhstan. Such information can be useful for formulating road safety strategies, taking into account the popularity of specific brands and their involvement in road accidents.


Frequency of accidents depending on the brands
# of Road Accident
2000
1750
1500
1250
1000
750
500
250
0
Toyota
Volkswagen
Mitsubishi
ВАЗ
Audi
Nissan
Другие
Mercedes-Benz
Subaru
Honda
Lexus
Mazda
BMW
ГАЗ
Opel
Hyundai
Suzuki
Daewoo
Kia
Lada
Ford
Лада
Mercedes
Москвич
Renault
УАЗ
Chevrolet
КАМАЗ
Daimler-Benz
Daimler Chrysler
Ssang Yong
Volvo
Skoda
Infiniti
Brand

Data analysis - Color

```
top_colors = df['Color'].value_counts().nlargest(10).index
df_filtered = df[df['Color'].isin(top_colors)]
plt.figure(figsize=(12, 8))
sns.countplot(data=df_filtered, x='Color', palette='coolwarm', order=top_colors)
plt.title('Frequency of Road Accidents by Top Colors')
plt.xlabel('Color')
plt.ylabel('# of Road Accident')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

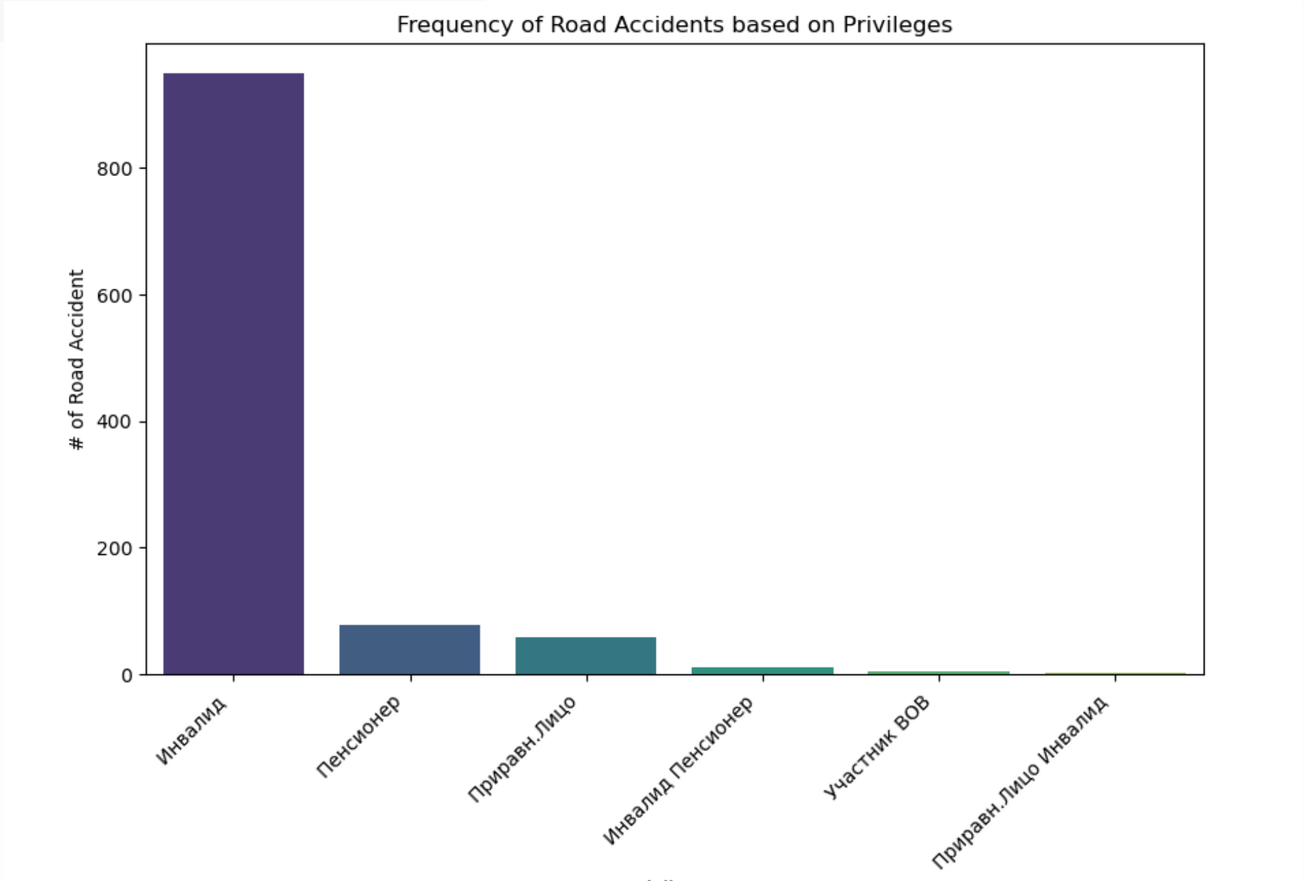


When analyzing car colors, the obvious and most basic colors, such as white, silver and black, are used more often than other colors. This may be due to their popularity among car owners and car manufacturers, as well as their better visibility on the road, especially in low light or bad weather conditions.

Data analysis - Privileges

```
privileged_accidents = df[df['Privileges'] != 'No privileges']
plt.figure(figsize=(10, 6))
sns.countplot(data=privileged_accidents, x='Privileges', palette='viridis')
plt.xticks(rotation=45, ha='right')
plt.xlabel('Privileges')
plt.ylabel('# of Road Accident')
plt.title('Frequency of Road Accidents based on Privileges')
plt.show()
```

People with privileges were involved in road accidents 1,101 times, and more than 80 of them were pensioners. These findings highlight the importance of studying the profiles of drivers with privileges and taking steps to improve their safety on the roads. Consideration of age and other factors can help identify driving habits and develop appropriate programs and educational campaigns to improve road safety among this group.



Data analysis - Descriptive statistics

```
selected_columns = ['Age', 'Driving_experience', 'Insurance_premium', 'Loss_amount']
df[selected_columns].describe()
```

	Age	Driving_experience	Insurance_premium	Loss_amount
count	9761.000000	9761.000000	9761.000000	9.761000e+03
mean	47.388280	18.084008	11321.710480	3.648240e+05
std	12.774785	8.758305	5781.744771	5.301417e+04
min	24.000000	0.000000	66.000000	1.572100e+04
25%	37.000000	12.000000	6852.000000	3.648240e+05
50%	46.000000	18.000000	10320.000000	3.648240e+05
75%	56.000000	20.000000	15720.000000	3.648240e+05
max	91.000000	88.000000	59669.000000	2.525000e+06

Thanks

An analysis of data on road traffic accidents (RTIs) in Kazakhstan has identified key factors influencing road safety. Retired people and those with benefits make up a significant portion of those involved in road accidents, which requires additional support programs and training. Car brands such as Toyota, Volkswagen and Mitsubishi are more likely to be involved in accidents, and basic colors such as white, silver and black dominate the statistics. Given these factors, it is necessary to develop and implement effective measures to improve road safety, including driver education, prevention campaigns and infrastructure improvements.

Garayev Emirdin