Faculty of Economic Sciences

Project Proposal

Moscow
2025

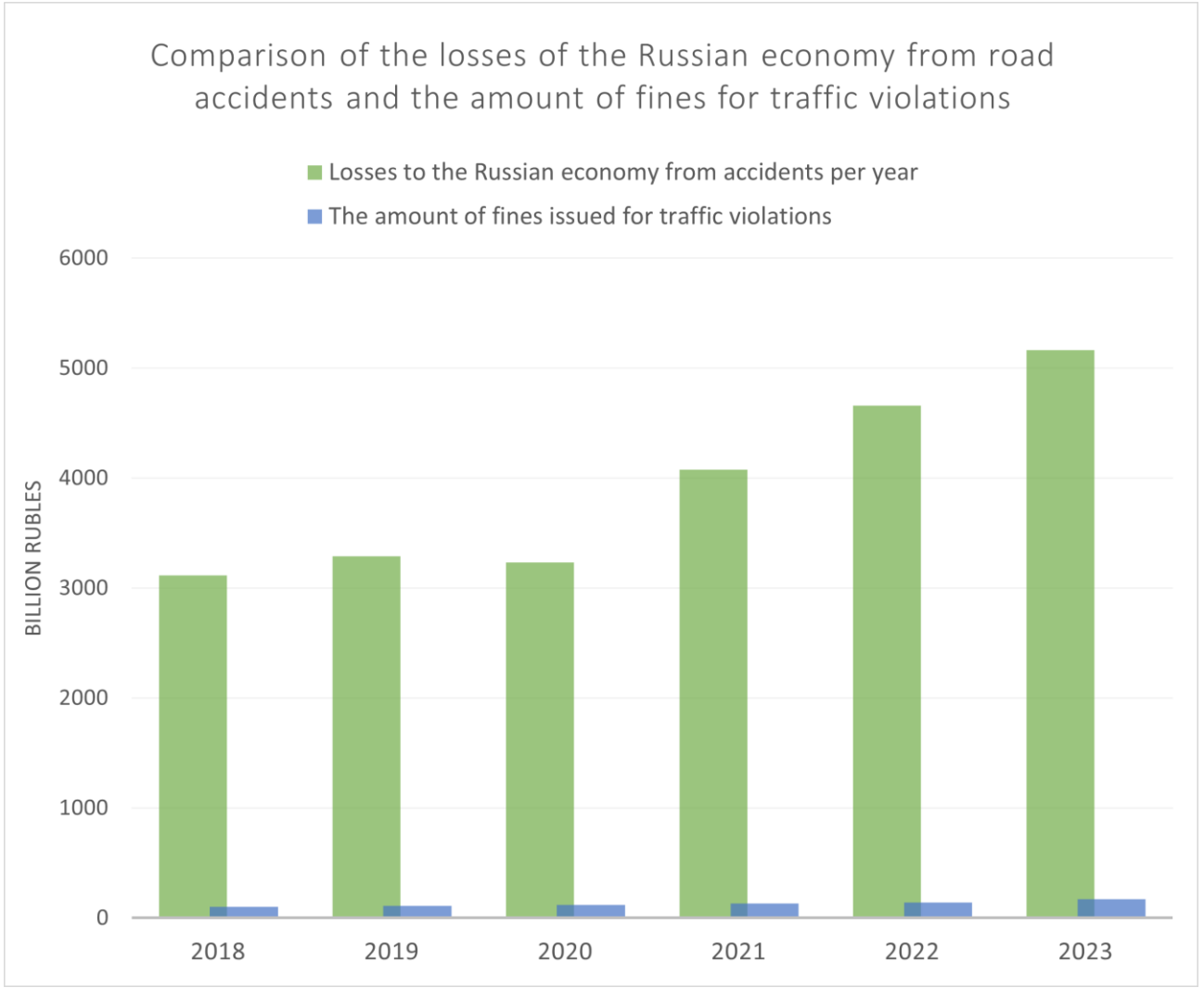
Research and Modeling of Accident Frequency and Severity on Toll Highways in Russia

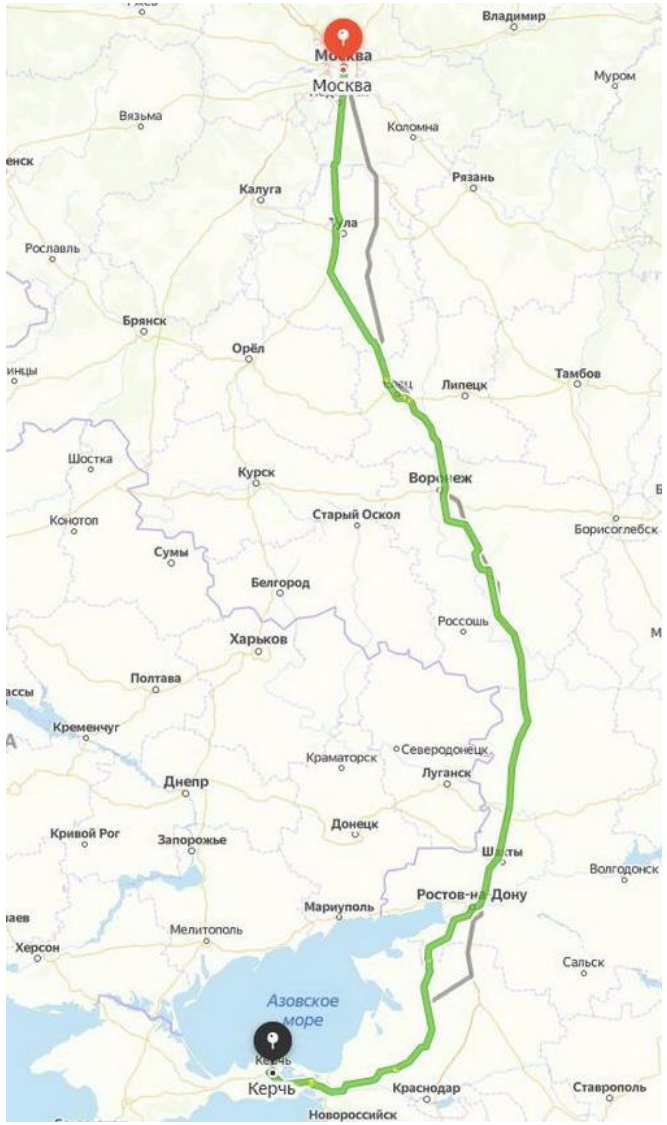
Author: Anastasia Kolesnikova,
group BEC216

Scientific supervisor: Ivan Stankevich,
Assistant Professor, Department of Applied Economics

Relevance of the Study

- Mortality and injuries resulting from road traffic accidents lead to economic losses amounting to 2% to 5% of a country's GDP.
- On average, 3% of global GDP is lost due to fatalities and injuries caused by road traffic accidents.
- The issue of road safety on toll roads and the economic benefits of investing in the development of the toll road system in Russia has not been studied comprehensively





Research Objectives and Tasks

Research question:

«How does road type (toll vs. free) affect the frequency and severity of traffic accidents in Russia?»

Research tasks:

- ❑ Collect, preprocess, and analyze traffic accident data for toll and free roads in Russia.
- ❑ Test the hypotheses using econometric methods and machine learning techniques.
- ❑ Identify the key factors influencing accident rates and compare them for toll and free roads.
- ❑ Identify "hot spots" and analyze whether key accident factors change in dangerous zones.
- ❑ Build a model to predict the likelihood of accidents in dangerous zones on toll and free roads.

Literature Review and Theory

- The introduction of tolls may lead to an increase in the number of accidents, as drivers start using alternative, less safe roads (Alfredo Marvão Pereira et al. (2020)

- The accident rate on toll roads is significantly influenced by factors such as the number of lanes, shoulder width, the presence of high-risk areas, and the distance to rest areas (Kristianto Kristianto, Carunia M. Firdausy, Najid Najid, Toni Hartono Bagio, 2024)

- The cost of preventing accidents is higher than the revenue from toll collection, while reducing toll rates could increase traffic on toll roads, potentially decreasing the number of accidents by redistributing traffic from less safe routes. (Zapico Emma, Baños-Pino José F., Mayor Matías, 2024)

Research Hypotheses

- **Hypothesis 1.1:**
«The average number of traffic accidents per month on toll roads is lower than on alternative free roads»

- **Hypothesis 1.2:**
«The factors influencing the number of traffic accidents differ for toll and free roads»

- **Hypothesis 2.1:**
«The probability of a more severe outcome of traffic accidents is lower on toll roads than on alternative free roads»

- **Hypothesis 2.2:**
«The factors influencing the severity of traffic accidents differ for toll and free roads»

Research Data

- **Data source:** Official website of the State Traffic Safety Inspectorate of the Ministry of Internal Affairs of Russia
- **Data description:** The statistics include data from all regions of Russia, covering the period from the beginning of 2015 to mid-2024
- **Number of observations:** 1,400,915 (1.4 million) accidents

Data Description

ОБЩИЕ СВЕДЕНИЯ

Дата	**29.01.2021**	Время	**16:30**
Широта	**53.99223**	Долгота	**35.08044**
Номер ДТП	**290003692**	Вид ДТП	**Съезд с дороги**

Схема

Адрес	**Думиничский район**
Дорога	**М-3 Украина Москва - Калуга - Брянск - граница с Украиной (основное направление), 268км 500м**
Значение дороги	**Федеральная (дорога федерального значения)**
Категория дороги	**4**
Категория улицы	**Вне НП**

ДОРОЖНЫЕ УСЛОВИЯ

Объекты УДС на месте ДТП	**Перегон (нет объектов на месте ДТП)**
Объекты УДС вблизи места ДТП	**Отсутствие в непосредственной близости объектов УДС и объектов притяжения**
Недостатки транспортно-эксплуатационного содержания улично-дорожной сети	**Не установлены**
Факторы, оказывающие влияние на режим движения	**Сведения отсутствуют**
Состояние погоды	**Пасмурно** **Снегопад**
Состояние проезжей части	**Обработанное противогололедными материалами**
Освещение	**Светлое время суток**
Изменения в режиме движения	**Режим движения не изменялся**

Data Description

УЧАСТНИКИ

Количество ТС **1**	Число участников **1**	Число погибших **0**	Число раненых **1**

ТС 1

Сведения об оставлении места	**Осталось на месте ДТП**	Расположение руля, тип привода	**Полноприводные**
Тип ТС	**D-класс (средний) до 4,6 м**	Год выпуска	**2008**
Марка/модель ТС	**HYUNDAI Прочие модели Hyundai**	Форма собственности	**Физические лица**
Цвет	**Черный**		
Места повреждения			
Технические неисправности	**Технические неисправности отсутствуют**		

УЧАСТНИК 1

Категория участника	**Водитель**	Использовался ли ремень	**Да**
Транспортное средство	**1**	Тип детского удерживающего устройства	
Сведения об оставлении места ДТП	**Нет (не скрывался)**	Степень опьянения	
Пол	**Женский**	Водительский стаж	**6**
Степень тяжести последствий	**Получил телесные повреждения с показанием к лечению в медицинских организациях (кроме разовой медицинской помощи)**		
Непосредственные нарушения ПДД	**Несоответствие скорости конкретным условиям движения**		
Сопутствующие нарушения ПДД	**Нет нарушений**		

- Panel data, where the observation is represented by a road section × month, indicating the number of accidents that occurred during that period (about 20-30 thousand observations)
- Time-location data, where each observation is an accident with information on time, geographical coordinates, and several other features (about 1 million observations)

Methods

For testing hypothesis (1.1) and (1.2):

- Panel **regression model** with FE:

$$Y_{it} = \beta_0 + \beta_1 Toll_i + \beta_2 X_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

- FE **Negative Binomial Regression**:

$$\ln Y_{it} = \beta_0 + \beta_1 Toll_i + \beta_2 X_{it} + \alpha_i + \gamma_t + \theta + \varepsilon_{it}$$

- Machine learning methods such as **Random Forest Regressor** and **Gradient Boosting** (using MAE/MSE metrics) with SHAP analysis or feature importance

For testing hypothesis (2.1) and (2.2):

- Multinomial Logit (MNL) regression:

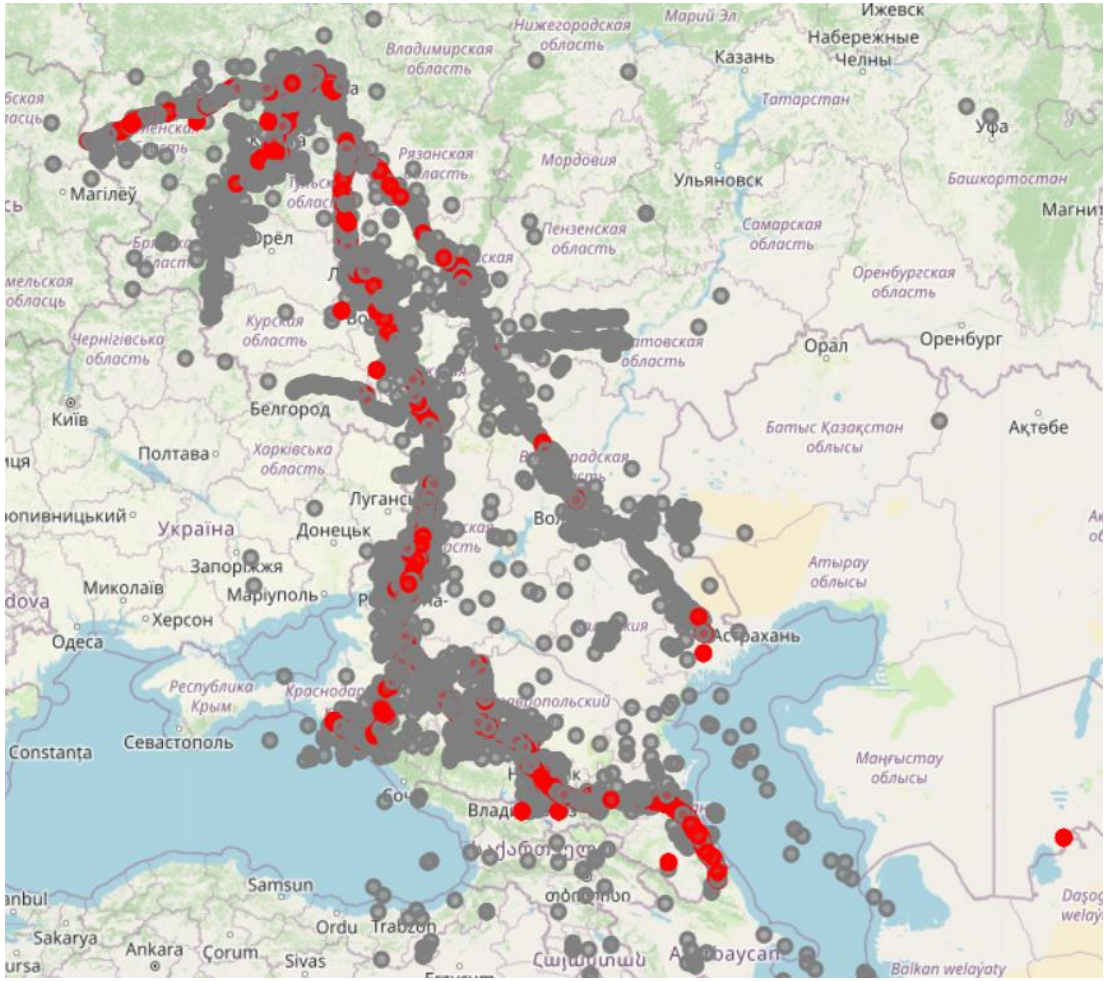
$$P(Y_i = k) = \frac{e^{\beta_k X_i}}{1 + \sum_{j=1}^{K} e^{\beta_j X_i}}$$

- Machine learning methods such as **Random Forest Regressor** and **Gradient Boosting** (using Accuracy and F1-score metrics) with SHAP analysis or feature importance

Methods

The key factors influencing accidents may differ from the overall sample on high-accident sections, so the next step will be to identify "hot spots"

- **DBSCAN** clustering algorithm
- **Predicting the probability of accidents** on these dangerous sections of different types of roads (toll vs. free) using **negative sampling**

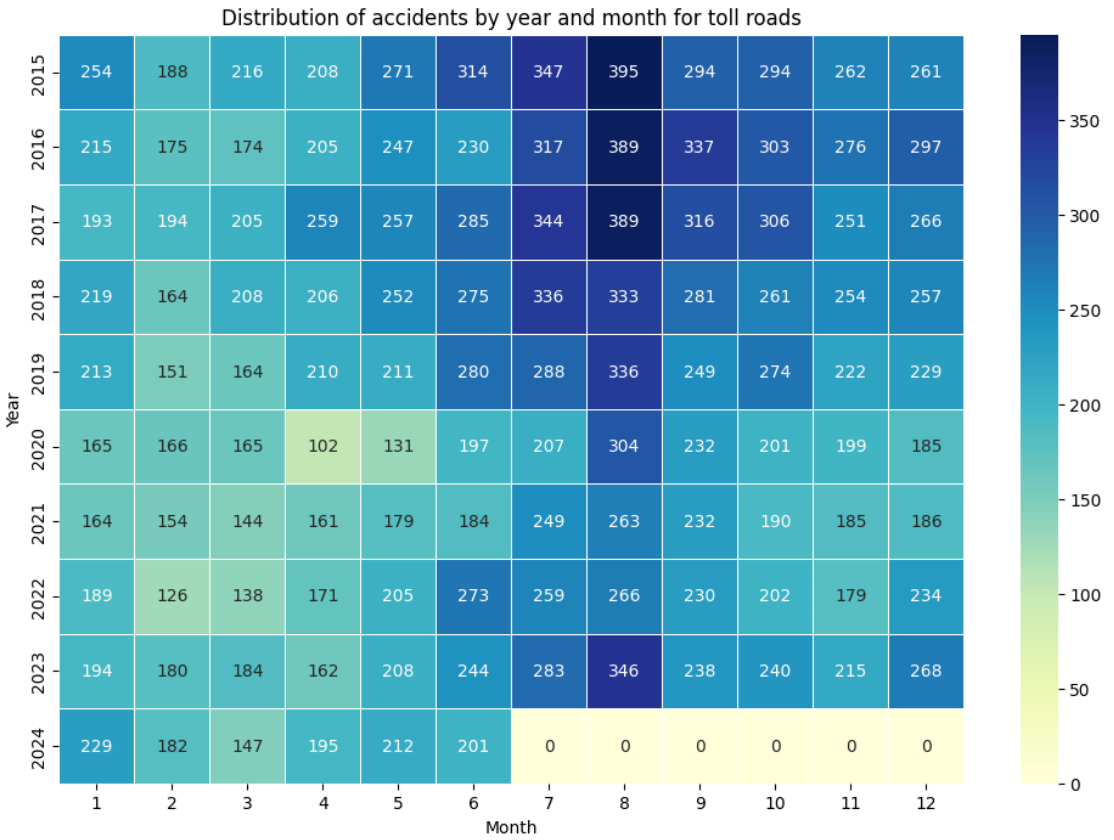


Results Anticipated

- The number of accidents on toll roads is indeed lower compared to free roads, but the severity of outcomes may be higher due to higher speed limits.
- The key factors influencing road safety include infrastructure quality, weather conditions, lighting, and traffic density.
- The results of the study will make it possible to offer recommendations for improving safety on both toll and free highways.

Appendix

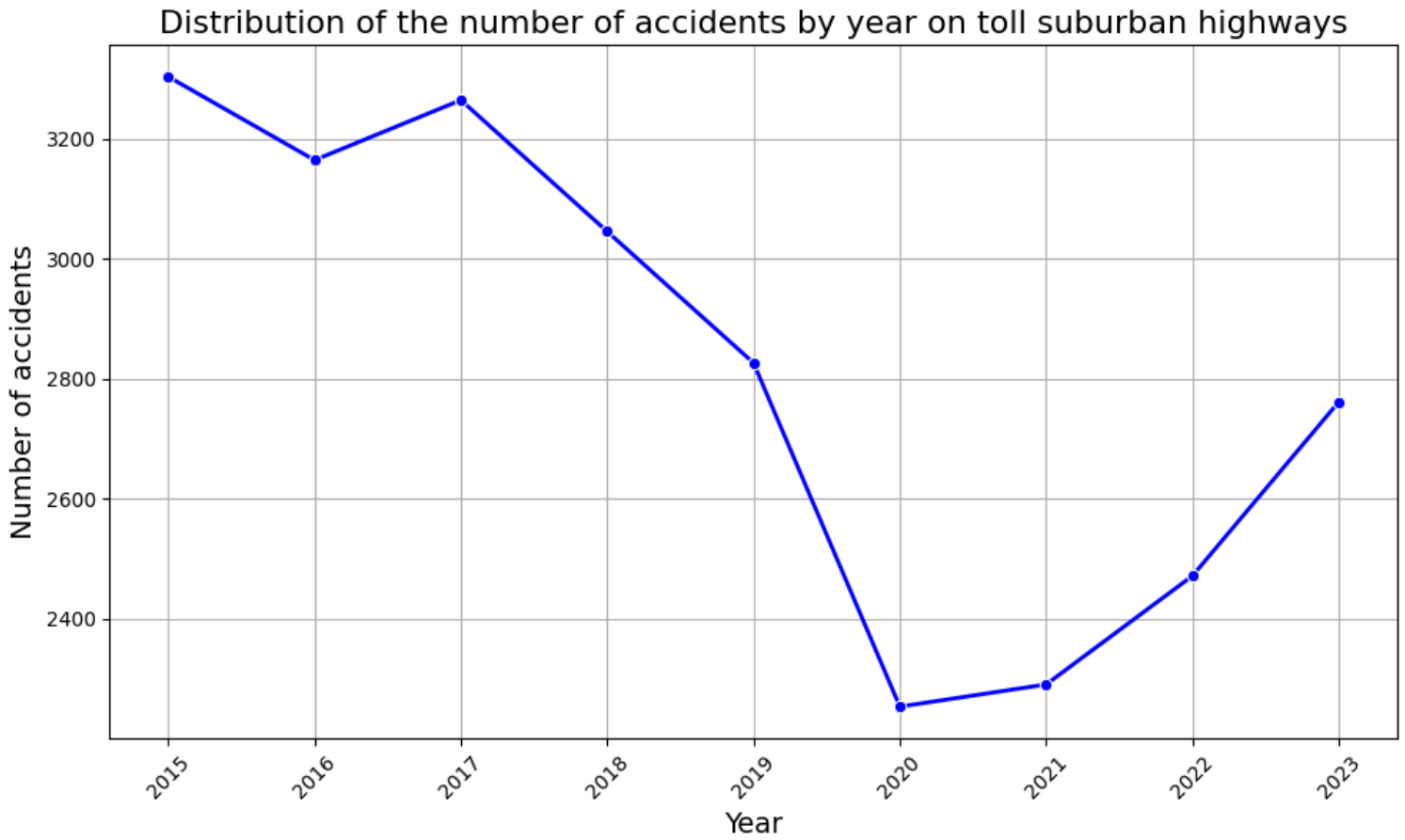

Distribution of accidents by year and month for free roads
Year
Month
2015 3350 2468 2516 2772 3974 4589 4854 5133 4494 4402 3803 4376
2016 3305 2524 2377 2674 3598 4067 5083 5317 4564 4335 3907 4102
2017 3006 2558 2241 2762 3498 3774 4735 4921 4340 4157 3709 3874
2018 3019 2067 2291 2428 3127 3857 4611 4695 4173 4011 3845 3928
2019 3054 2281 2416 2540 3249 3994 4340 4468 4017 3862 3633 3942
2020 3142 2655 2297 1895 2798 3371 4234 4373 3730 3516 3352 3409
2021 2717 2105 2100 2153 2924 3433 4073 3951 3322 3415 3252 3456
2022 2774 1917 2198 2164 2810 3574 4153 4132 3396 3401 3225 3505
2023 2682 2170 2123 2544 3058 3472 4004 4219 3686 3630 3290 3379
2024 2704 2168 1947 2362 2727 3310 0 0 0 0 0 0
1 2 3 4 5 6 7 8 9 10 11 12
5000 4000 3000 2000 1000 0


Distribution of accidents by year and month for toll roads
Year
Month
2015 254 188 216 208 271 314 347 395 294 294 262 261
2016 215 175 174 205 247 230 317 389 337 303 276 297
2017 193 194 205 259 257 285 344 389 316 306 251 266
2018 219 164 208 206 252 275 336 333 281 261 254 257
2019 213 151 164 210 211 280 288 336 249 274 222 229
2020 165 166 165 102 131 197 207 304 232 201 199 185
2021 164 154 144 161 179 184 249 263 232 190 185 186
2022 189 126 138 171 205 273 259 266 230 202 179 234
2023 194 180 184 162 208 244 283 346 238 240 215 268
2024 229 182 147 195 212 201 0 0 0 0 0 0
1 2 3 4 5 6 7 8 9 10 11 12
350 300 250 200 150 100 50 0

Appendix


Distribution of the number of accidents by year on free suburban highways
Number of accidents
46000
44000
42000
40000
38000
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year

Appendix


Distribution of the number of accidents by year on toll suburban highways
Number of accidents
3200
3000
2800
2600
2400
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year


ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •