

Предсказание количества ДТП

Громяков Никита, 332 г



Вступление

Дорожно-транспортные происшествия (ДТП) являются одной из основных причин травматизма и смертности на дорогах. Частота аварий зависит от множества факторов: времени суток, дня недели, опыта водителя, типа транспортного средства и особенностей региона. Цель проекта — выявить закономерности, позволяющие прогнозировать число ДТП и использовать эти знания для профилактики.

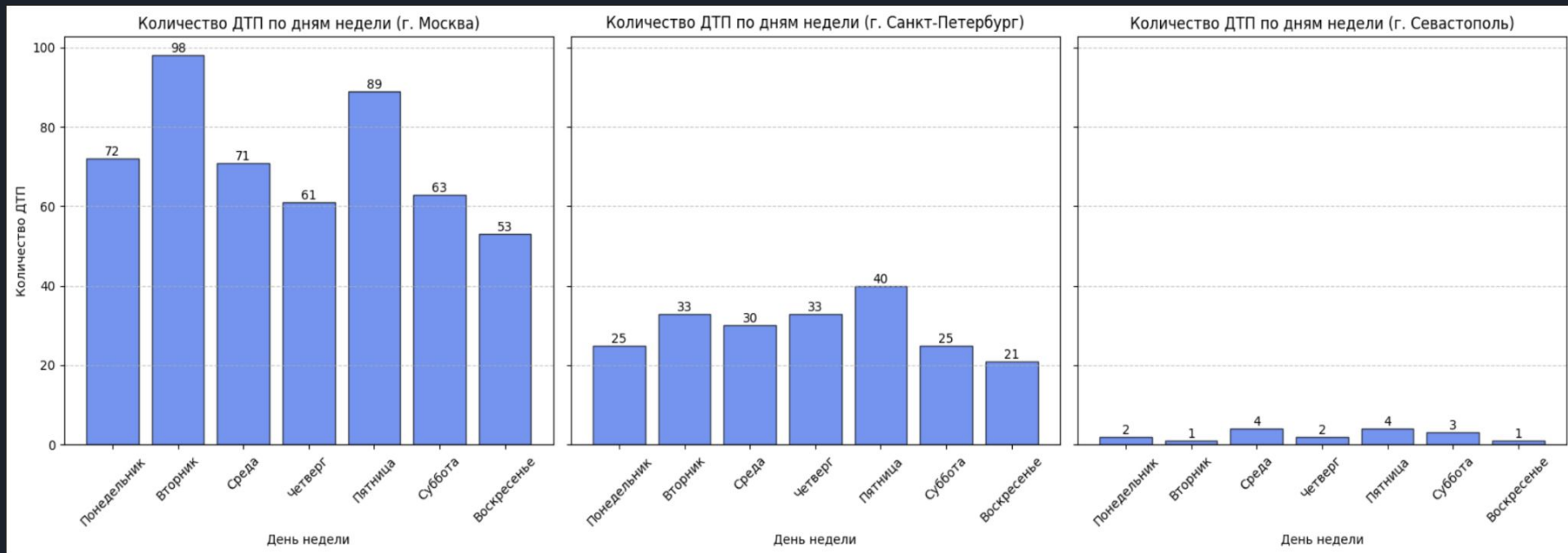
Задача: С помощью линейной регрессии получить модель, предсказывающую число ДТП в зависимости от дня недели, времени суток и других параметров.

Цель: Получить интерпретируемую модель, отражающую влияние показателей на частоту ДТП.

Исходные данные

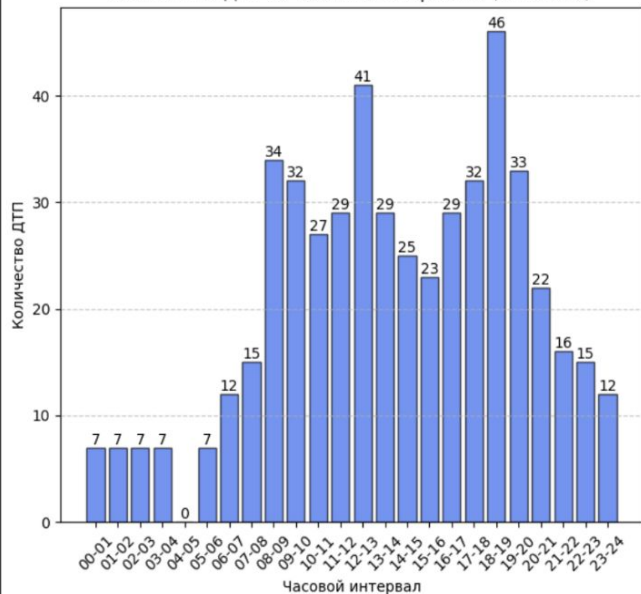
Наименование показателя	№ строки	гор. Москва					
		ДТП		Погибло		Ранено	
		абс.	± % к АППГ	абс.	± % к АППГ	абс.	± % к АППГ
А	Б	1	2	3	4	5	6
Дорожно-транспортные происшествия	1	526	-4,0	28	100,0	604	-3,7
–из них–							
–по дням недели–							
произошедшие по понедельникам	2	56	-25,3	2	100,0	65	-24,4
произошедшие по вторникам	3	70	-9,1	1	стаб.	78	-4,9
произошедшие по средам	4	80	-5,9	7	250,0	89	-3,3
произошедшие по четвергам	5	72	-22,6	2	100,0	79	-28,2
произошедшие по пятницам	6	90	4,7	7	16,7	101	9,8
произошедшие по субботам	7	85	19,7	7	75,0	104	26,8
произошедшие по воскресеньям	8	73	19,7	2	100,0	88	6,0
–по времени суток–							
произошедшие с 00 по 01 часов	9	9	12,5	0	-100,0	11	22,2
произошедшие с 01 по 02 часов	10	8	-33,3	3	200,0	12	-20,0
произошедшие с 02 по 03 часов	11	6	20,0	1	100,0	5	-16,7
произошедшие с 03 по 04 часов	12	5	стаб.	0	-100,0	10	42,9
произошедшие с 04 по 05 часов	13	6	100,0	3	100,0	10	150,0
произошедшие с 05 по 06 часов	14	3	-50,0	0	-100,0	4	-33,3
произошедшие с 06 по 07 часов	15	10	100,0	1	100,0	14	180,0
произошедшие с 07 по 08 часов	16	19	-9,5	1	100,0	29	7,4
произошедшие с 08 по 09 часов	17	30	-11,8	2	100,0	32	-15,8
произошедшие с 09 по 10 часов	18	23	9,5	0	-100,0	29	3,6
произошедшие с 10 по 11 часов	19	27	стаб.	2	100,0	28	-12,5

Анализ данных ДТП по дням недели

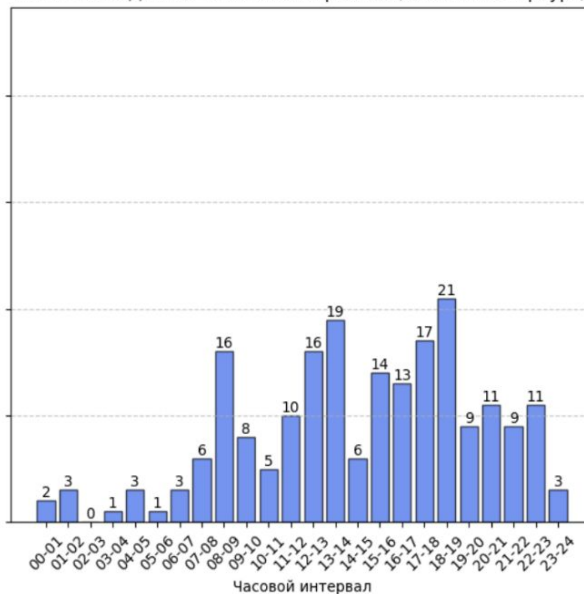


Анализ данных ДТП по часовым интервалам

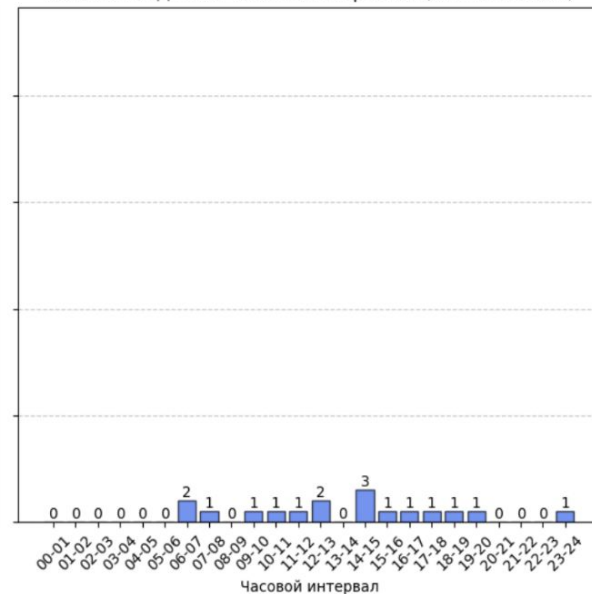
Количество ДТП по часовым интервалам (г. Москва)



Количество ДТП по часовым интервалам (г. Санкт-Петербург)



Количество ДТП по часовым интервалам (г. Севастополь)



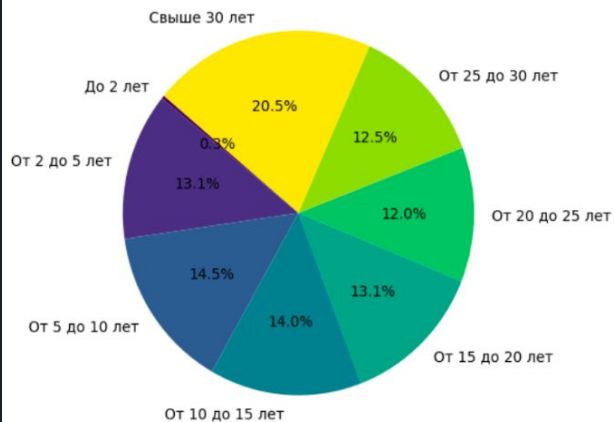
Анализ данных ДТП по часовым интервалам



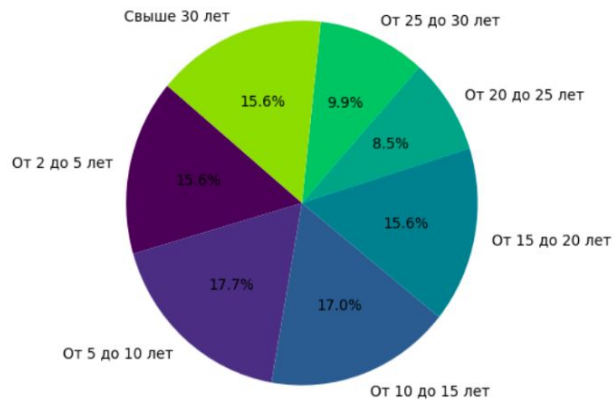
Анализ данных

Распределение ДТП по стажу

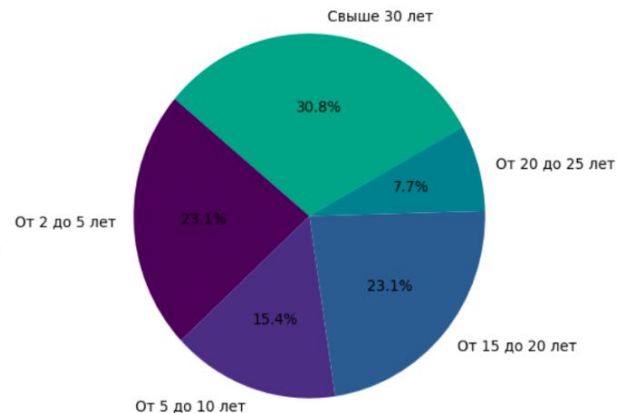
Распределение ДТП по стажу водителей (г. Москва)



Распределение ДТП по стажу водителей (г. Санкт-Петербург)



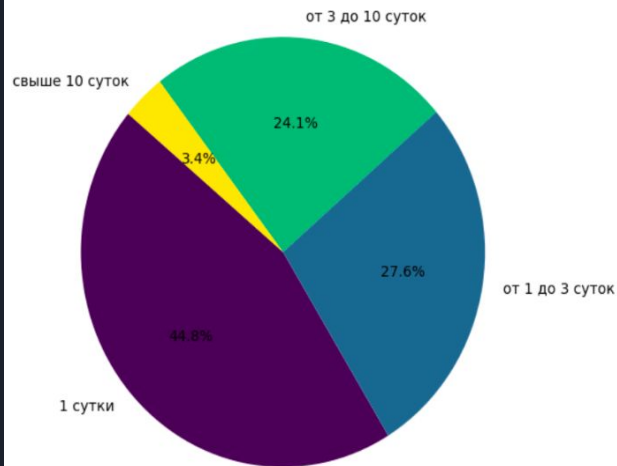
Распределение ДТП по стажу водителей (г. Севастополь)



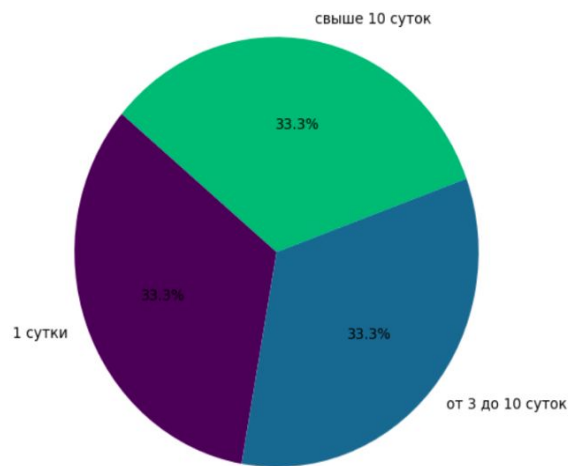
Анализ данных

Задержания в срок, скрывшихся с места ДТП

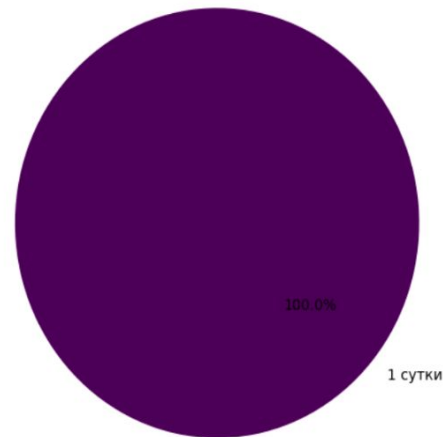
Задержаны в срок, скрывшиеся с места ДТП (г. Москва)



Задержаны в срок, скрывшиеся с места ДТП (г. Санкт-Петербург)



Задержаны в срок, скрывшиеся с места ДТП (г. Севастополь)



Обучение модели

Линейная регрессия

Две модели: по дням и по часовым интервалам

Признаки:

год, месяц, сезон, день недели,
выходной/будний день, час дня, время дня
(утро/день/вечер/ночь)

Результаты:

Model for days of week:

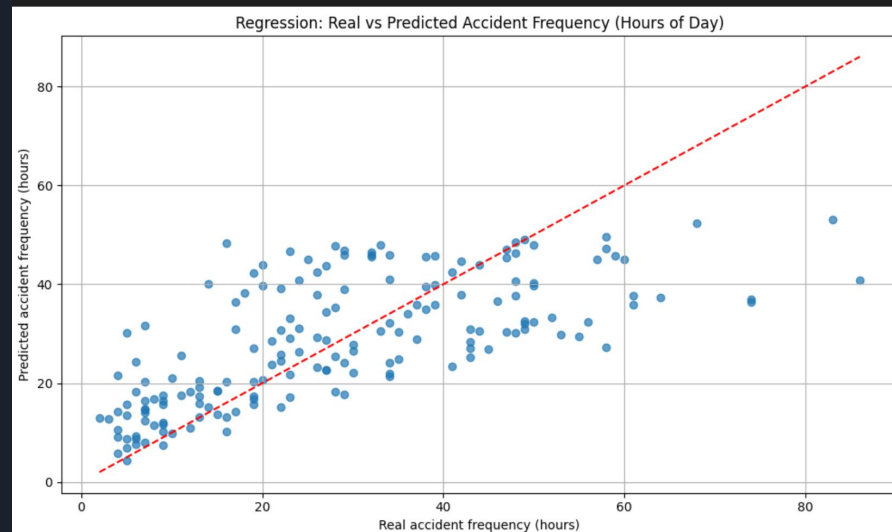
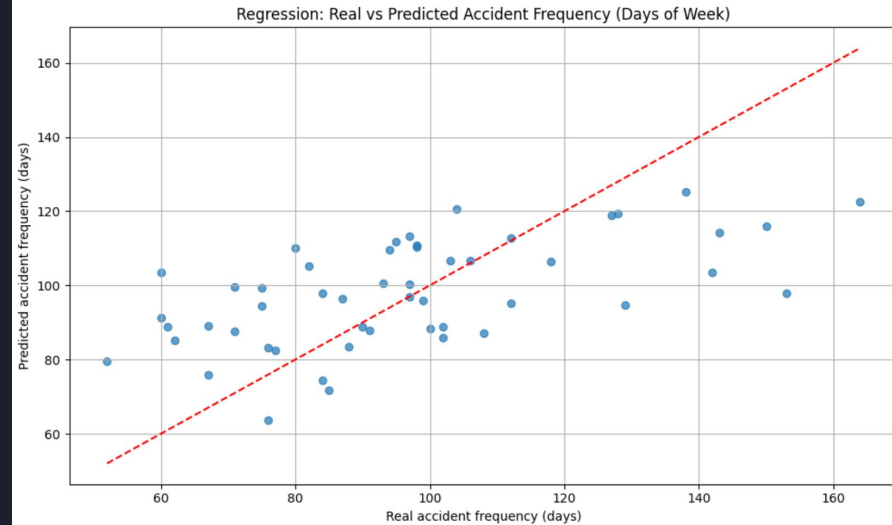
R2 score: 0.3477

MSE: 439.404495

Model for hours of day:

R2 score: 0.4777

MSE: 173.190093



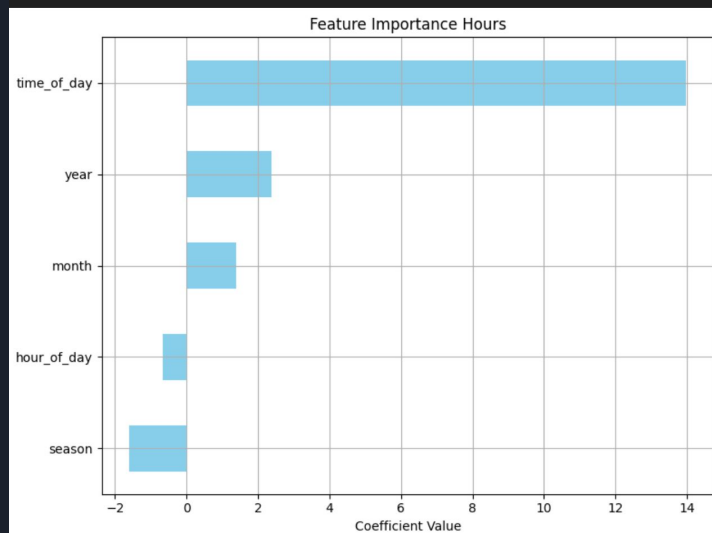
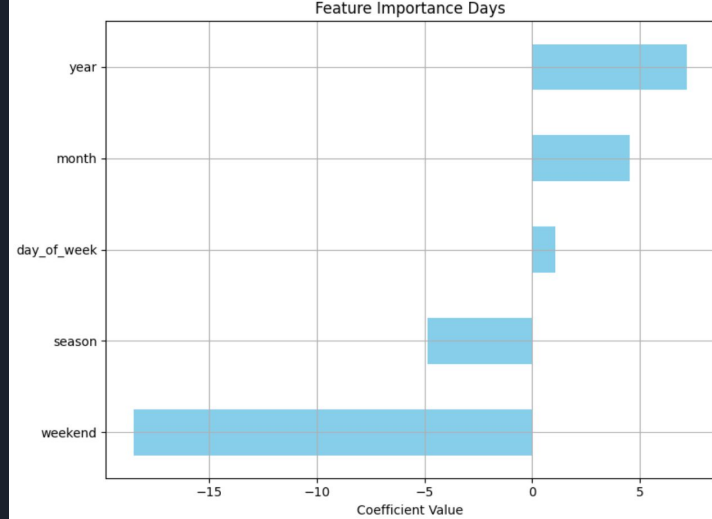
Обучение модели

Линейная регрессия

Важные параметры:

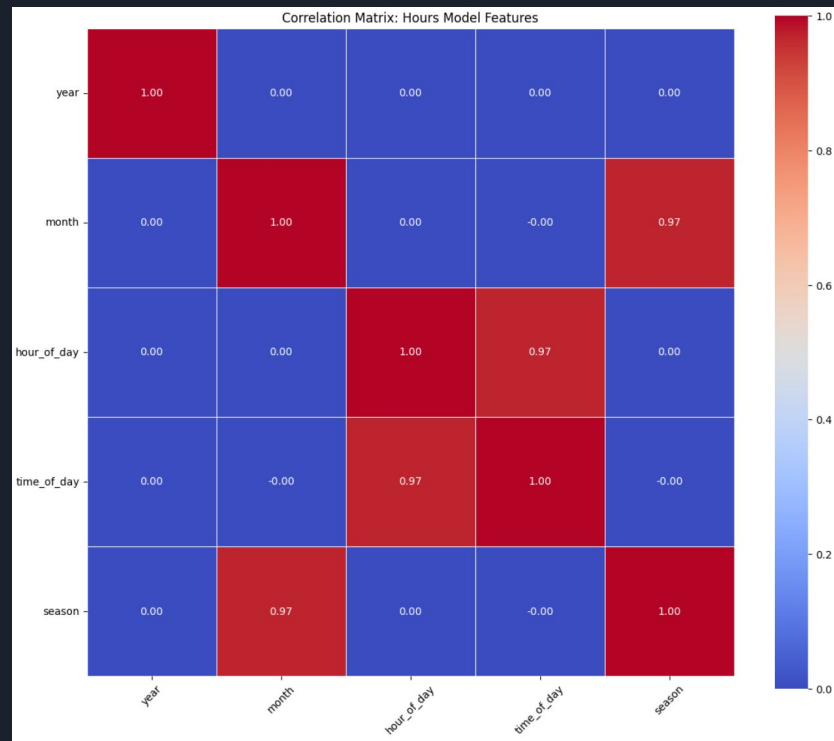
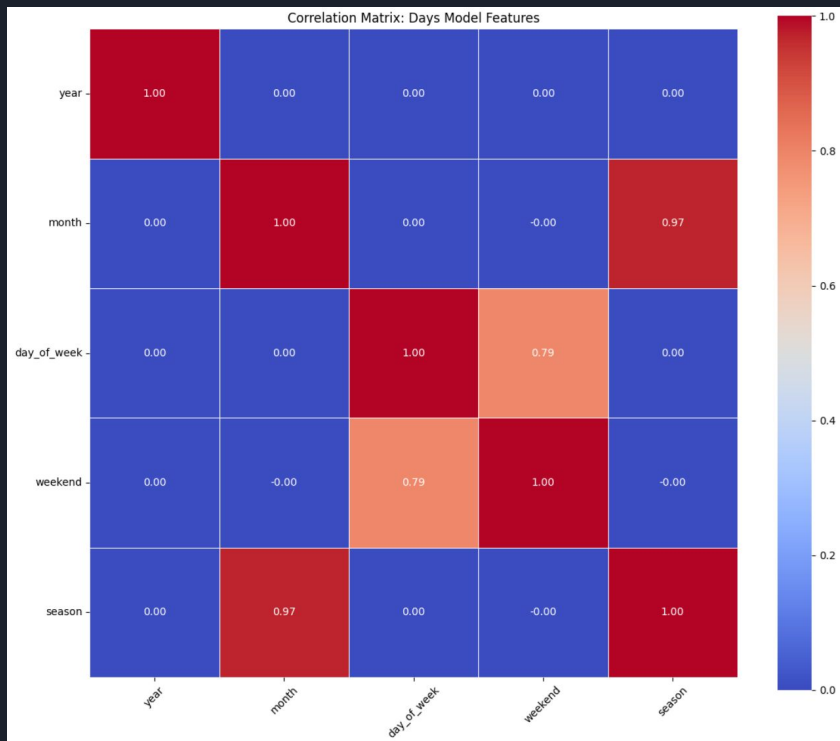
Модель по дням: выходной/будний день

Модель по часовым интервалам: время дня
(утро/день/вечер/ночь)



Обучение модели

Линейная регрессия



Линейная регрессия

One hot encoding

Model for days of week:

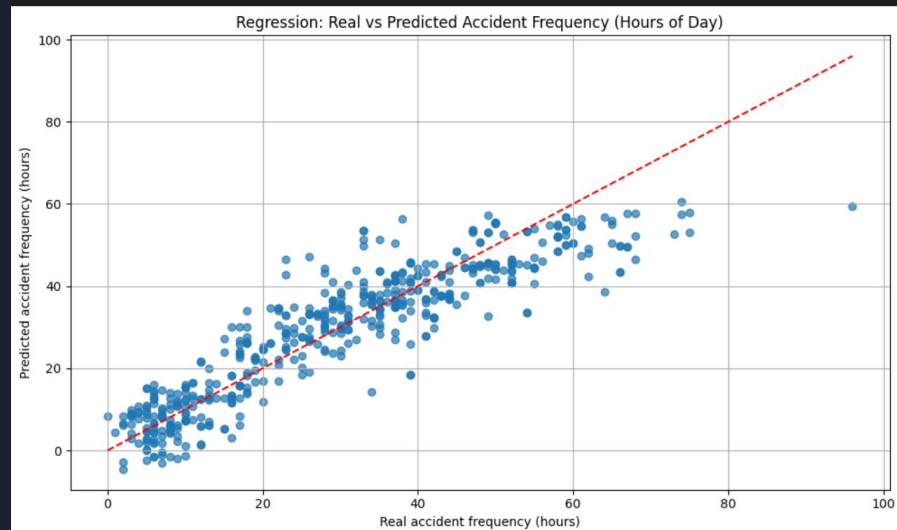
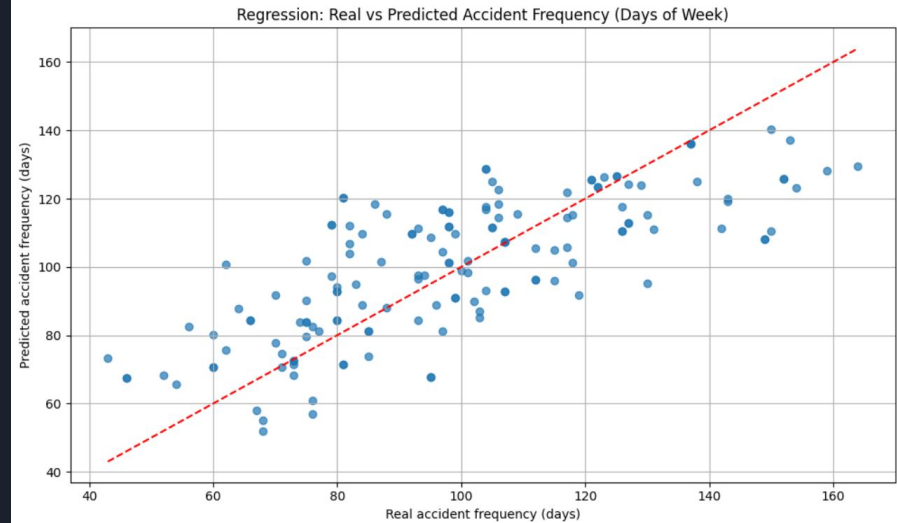
R2 score: 0.5532

MSE: 308.182778

Model for hours of day:

R2 score: 0.8230

MSE: 58.835230



Обучение модели RandomForestRegressor

Model for days of week:

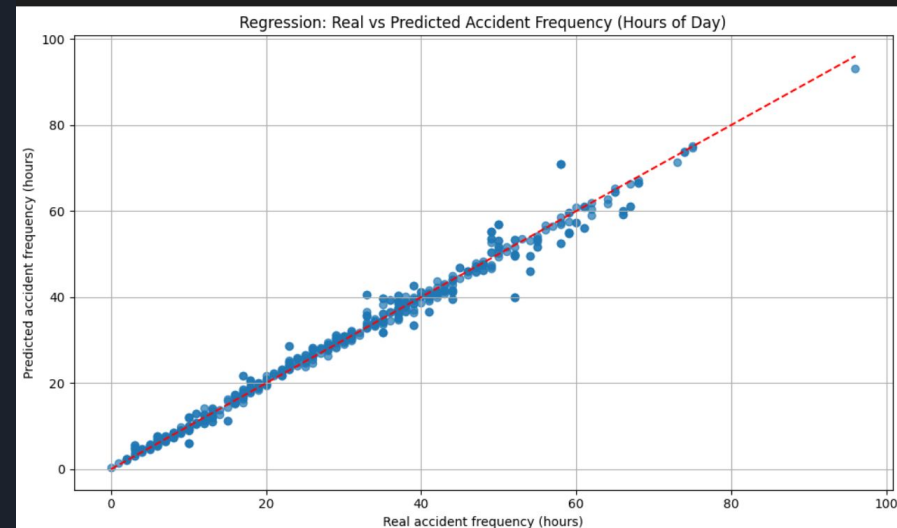
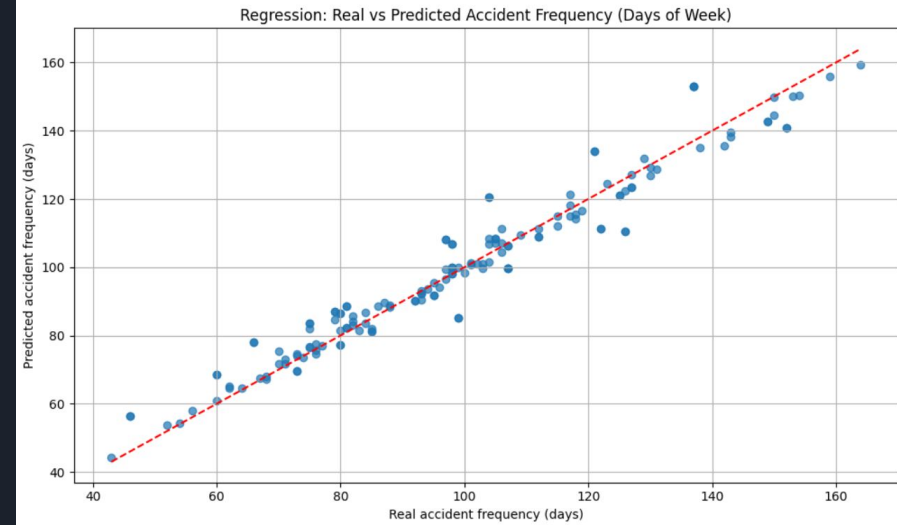
R2 score: 0.9470

MSE: 36.527400

Model for hours of day:

R2 score: 0.9825

MSE: 5.832081



Обучение модели RandomForestRegressor

Model for days of week:

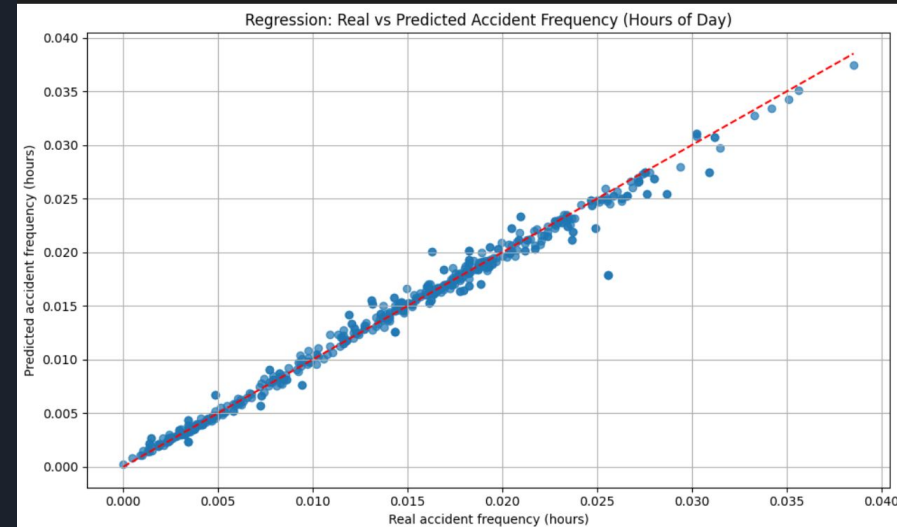
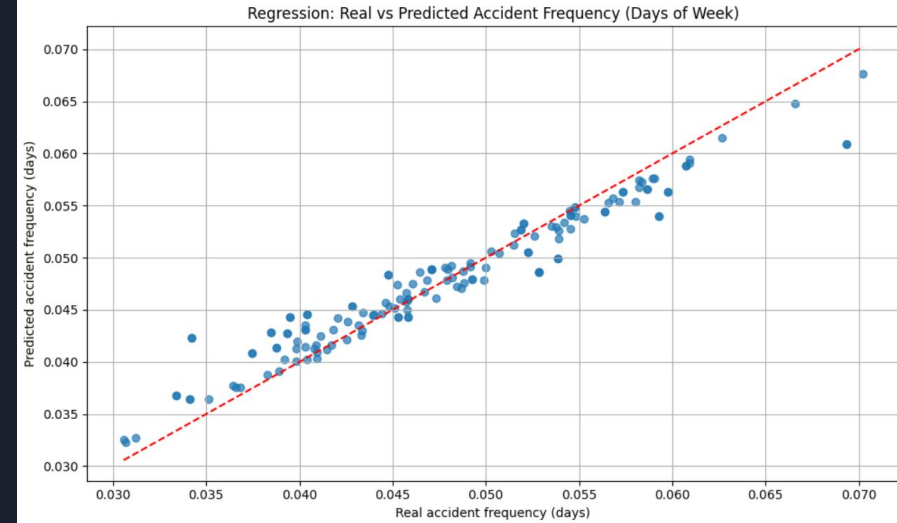
R2 score: 0.9155

MSE: 0.000006

Model for hours of day:

R2 score: 0.9838

MSE: 0.000001



Результат

