

# Трафик, аварийность и ремонтные работы на участках дорог М-8 (Москва – Северодвинск) и Р-21 (Санкт-Петербург – Борисоглебск)

Лицензия: Продвинутого режима

Доступ: Продвинутый

Дата обновления: 09.09.2021

Теги:

транспорт, дороги, ДТП, аварийность

## Описание

Набор данных содержит сведения, характеризующие движение на различных участках автомобильных дорог М-8 (E115) «Холмогоры» и Р-21 (E105) «Кола» за 2015–2021 гг. Впервые данные были опубликованы и использованы для построения ML-моделей в рамках хакатона [«Emergency DataHack 2021»](#).

Единица наблюдения в датасете – километр указанных дорог за определенный период времени (в том числе с точностью до часа и выше). Данные представлены в трех таблицах, которые могут быть объединены между собой по дате и километру (участку) дороги. По каждой единице наблюдения в основной таблице представлены следующие атрибуты: дата измерения, направление движения, значение общей интенсивности на дороге, загрузка полосы в процентах, средняя скорость на данном участке, координаты широты и долготы и т.д. В дополнительных таблицах содержатся атрибуты: километр дороги, отнесенный к аварийно-опасному участку в конкретный год, точки начала и окончания этого участка, его протяженность, географическая принадлежность к населенному пункту, характер стабильности местоположения участка концентрации ДТП, а также атрибуты, посвященные ремонтным работам: описание ремонтных работ, период их проведения, протяженность, затраты на проведение ремонтных работ и т.д.

Набор доступен для работы в формате: CSV(кодировка: «UTF-8», разделитель: «;»).

## Характеристики датасета

### Использование

Лицензия: [Продвинутого режима](#)

Доступ: Продвинутый

### Предоставление данных

Формат:

CSV

### Характеристики набора

Временной период: 01.01.2015 – 14.04.2021

Количество наблюдений: 0

### Обновления

Размещено: 09.09.2021

Обновлено: 09.09.2021

Частота обновления: не указано

Дата следующего обновления: не указано

### Происхождение

Источники данных: Федеральное дорожное агентство

Ответственные за набор данных: Цыганков Максим Валерьевич, Журавлев Данила Олегович

# Структура набора данных

## Характеристики интенсивности движения (traffic)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
data_id	Представляет идентификационный номер (ID) конкретного измерения, произведенного на пункте учета интенсивности движения (ПУИД)	0			integer
station_id	ID ПУИД, с использованием которого было получено измерение (наблюдение)	0			integer
direction	Направление движения транспортного средства для данного наблюдения. Представлено двумя значениями: • forward – при движении от нулевого километра; • backward – при движении к нулевому километру	0			cat
lane_count	Общее количество полос на участке автомобильной дороги по отношению к расположению ПУИД. Принимает только целые значения от 1 до 6	0			integer
lane	Номер полосы, по которой осуществляется движение. В датасете представлена целыми значениями от 1 до 6. Значение этой переменной и переменной «direction» позволяет определить направление движения по каждой полосе	0			integer
volume	Общая интенсивность движения на дороге. Рассчитывается как общее количество транспортных средств, пересекших конкретное сечение автодороги за единицу времени по соответствующему направлению движения.	0	АВТ/Ч		numeric
occupancy	Значение загрузки полосы в процентах. Принимает значения от 0 до 100	0			numeric
speed	Средняя скорость на данном участке дороги, км/ч за время наблюдения	0	КМ/Ч	333	numeric
latitude	Координата широты для данного километра дороги в градусах	0	градус		numeric

longitude	Координата долготы для данного километра дороги в градусах.	0	градус		numeric
datetime	Данные о дате измерения в формате гггг-мм-дд чч:мм:сс	0	YYYY-MM-DD HH:MM:SS		date
road	ID автомобильной дороги. В датасете переменная принимает два значения: • 9 – М-8 (Е115); • 14 – Р-21 (Е105)	0			cat
road	Километр автомобильной дороги в соответствии с разметкой километража по направлению движения	0	КМ	008	integer
name	Расположение данного участка по километрам рассматриваемой трассы. Пример: "км 180+700", где 180 – номер километра, а 700 – метры от последнего пройденного километра	0			text

### Аварийно-опасные участки (crash\_parts)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
length	Протяженность аварийно-опасного участка в метрах. Тут же представлены значения для идентификации особенных случаев, в частности: 1 – транспортная развязка; 2 – пересечение; 3 – примыкание	0	КМ	008	integer
avuch_loc	Местоположение аварийного участка. Принимает два значения: 1 – в пределах населенного пункта; 2 – вне пределов населенного пункта	0			integer
datetime	Данные о дате измерения в формате гггг-мм-дд чч:мм:сс	0	YYYY-MM-DD		date
road_id	ID автомобильной дороги. В датасете переменная принимает два значения: • 9 – М-8 (Е115); • 14 – Р-21 (Е105)	0			integer
road_km	Километр автомобильной дороги в соответствии с разметкой километража по направлению движения	0	КМ	008	integer
avuch_start	Точка начала аварийного участка с точностью до метра.	0	КМ	008	numeric
avuch_end	Точка окончания аварийного участка с точностью до метра	0	КМ	008	numeric

stabchar_type	Стабильность местоположения участка концентрации ДТП. Принимает следующие три значения: 1 – стабильный; 2 – мигрирующий (подразумевается варьирующий характер частоты ДТП на данном участке при сравнении по годам); 3 – вновь возникший (подразумевает возникновение нескольких аварийно-опасных ситуаций в рамках одного года)	0			integer
planactiv_type	Код планируемых работ на данном участке. Принимает следующие четыре значения: 1 – реконструкция; 2 – капитальный ремонт; 3 – ремонт; 4 – содержание	0			integer
planactiv_descr	Подробное описание планируемого/планируемых работ на данном участке	0			text
planactiv_year	Планируемая дата окончания работ на данном участке	0	YYYY-MM-DD		date

## Ремонтные работы (repair)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
datetime	Данные о дате измерения в формате гggг-мм-дд чч:мм:сс	0	YYYY-MM-DD		date
road_id	ID автомобильной дороги. В датасете переменная принимает два значения: • 9 – М-8 (E115); • 14 – Р-21 (E105)	0			integer
road_km	Километр автомобильной дороги в соответствии с разметкой километража по направлению движения	0	КМ	008	integer
repair_id	Идентификационный номер проведенных ремонтных работ (в первоисточнике не содержится; присваивался самостоятельно при обработке датасета)	0			integer
repair_description	Наименование ремонтных работ согласно отчетным данным	0			text
repair_period	Период, в который проходили ремонтные работы	0			text
length	Протяженность отремонтированного участка в километрах	0	КМ	008	numreic

price	Стоимость ремонтных работ, тыс. руб.	0			numeric
direction	Направление движения транспортного средства по данному наблюдению. Представлена двумя значениями: • forward – при движении от нулевого километра; • backward – при движении к нулевому километру	0			cat

## Внесенные в набор данных изменения по сравнению с источниками

В таблице traffic.csv были выполнены следующие процедуры обработки данных.

- Для данных по автомобильной дороге М-8:

- заменено 47 899 (1,3% от выборки) нулевых значений заданного трафика на ненулевой расчетный трафик<sup>1</sup>;
- заменено 29 259 (0,79% от выборки) пропусков средней скорости на нули для нулевого трафика;
- удалена 16 471 (0,45% от выборки) запись с нулевыми значениями расчетной интенсивности трафика;
- удалено 25 528 (0,7% от выборки) записей с пропусками значений загрузки полосы;
- заменено 8479 (0,23% от выборки) ненулевых записей средней скорости при нулевом трафике на нулевые;
- удалено 6229 (0,17% от выборки) значений средней скорости, превышающей 200 км/ч;
- заменен 4901 (0,13% от выборки) аутлаер<sup>2</sup> значений трафика на значения расчетной интенсивности трафика;
- удалено 57 (0,002% от выборки) аутлаеров значений загрузки полосы, превышающих 100%;
- всего удалено 48 285 наблюдений (1,31% от первоначальной выборки).

- Для данных по автомобильной дороге Р-21:

- заменено 73 903 (1,98% от выборки) нулевых значения заданного трафика на ненулевые расчетного трафика<sup>1</sup>;

- удалено 4 (0,0001% от выборки) записи с пропусками значений загрузки полосы;
- заменено 42 065 (1,12% от выборки) ненулевых записей средней скорости при нулевом трафике на нулевые;
- удалено 7724 (0,21% от выборки) значения средней скорости, превышающей 200 км/ч;
- заменено 20 467 (0,55% от выборки) аутлаеров значений трафика на значения расчетной интенсивности трафика<sup>2</sup>;
- удалено 7 (0,0002% от выборки) аутлаеров значений трафика, превышающих порог в 10 000 ТС/ч;
- удалено 32 (0,0009% от выборки) аутлаера значений загрузки полосы, превышающих 100%;
- всего удалено 7767 наблюдений (0,21% от первоначальной выборки).

[\[i\]](#)

---

<sup>1</sup>Заданный трафик / интенсивность движения – автоматически рассчитанная АПУИД сумма всех

транспортных средств (имеет погрешность вычислений). Расчетный трафик / интенсивность движения – восстановлен на основании дополнительно предоставленных Росавтодором агрегированных данных АПУИД по количеству транспортных средств в разрезе их типов.

<sup>2</sup> Аутлаеры определялись на основе регрессионной модели с учетом величины распределения остатка (не больше трех стандартных отклонений).[i]



#### Источник 4. Объемы работ на 2015–2020 гг. по ремонту искусственных дорожных сооружений на действующей сети автомобильной дороги общего пользования федерального значения

Полное наименование источника данных	Объемы работ на 2015–2020 гг. по ремонту искусственных дорожных сооружений на действующей сети автомобильной дороги общего пользования федерального значения
Сокращенное наименование источника данных	Объемы работ по ремонту искусственных дорожных сооружений
Владелец источника данных	Федеральное дорожное агентство (Росавтодор)
Краткое описание источника данных	Источник данных представлен в виде таблицы с характеристиками по проводимым ремонтным работам на искусственных дорожных сооружениях на действующей сети автомобильной дороги общего пользования федерального значения, их виду, со значением километра, описанием ремонтных работ, периодом их проведения, протяженностью участка и затратами на проведение ремонтных работ
Методология и изменения методологии для источника данных	ОДМ 218.3.082-2016 «Методические рекомендации по назначению технологий и периодичности проведения работ по устройству слоев износа и защитных слоев дорожных покрытий».  ГОСТ Р 58862-2020 «Дороги автомобильные общего пользования. Ремонтные работы. Периодичность проведения»
Ссылки на методологию	<a href="https://docs.cntd.ru/document/456041198">https://docs.cntd.ru/document/456041198</a>  <a href="https://docs.cntd.ru/document/1200173464">https://docs.cntd.ru/document/1200173464</a>

#### Источник 1. Регистрационные данные автоматизированных пунктов учета интенсивности движения (АПУИД) за период 2016–2021 гг.

Полное наименование источника данных	Регистрационные данные автоматизированных пунктов учета интенсивности движения (АПУИД) за период 2016–2021 гг.
Сокращенное наименование источника данных	Регистрационные данные АПУИД за 2016–2021 гг.
Владелец источника данных	Федеральное дорожное агентство (Росавтодор)
Краткое описание источника данных	Данные по интенсивности движения являются массивом наблюдений, полученных с использованием отчетных данных из Формы представления данных учета интенсивности и состава движения с пунктов автоматизированного учета движения, с которой можно ознакомиться в ОДМ 218.2.032-2013 (Приложение Д). Указанная форма фиксирует общее количество автомобилей (всего и по различным спецификациям), которые к конкретному моменту проехали анализируемый участок конкретной дороги в единицу времени
Ссылка на источник данных	Прямая ссылка отсутствует, источник – информационная система ограниченного доступа
Методология и изменения методологии для источника данных	ОДМ 218.2.032-2013 «Методические рекомендации по учету движения транспортных средств на автомобильных дорогах»

Ссылки на методологию	<a href="https://rosavtodor.gov.ru/storage/app/media/uploaded-files/218.2.032-2013.pdf">https://rosavtodor.gov.ru/storage/app/media/uploaded-files/218.2.032-2013.pdf</a>
-----------------------	---

### Источник 3. Объемы работ на 2015–2020 гг. по содержанию действующей сети автомобильных дорог общего пользования федерального значения (устройству защитных слоев, слоев износа и поверхностной обработки)

Полное наименование источника данных	Объемы работ на 2015–2020 гг. по содержанию действующей сети автомобильных дорог общего пользования федерального значения (устройству защитных слоев, слоев износа и поверхностной обработки)
Сокращенное наименование источника данных	Объемы работ по содержанию действующей сети автомобильных дорог
Владелец источника данных	Федеральное дорожное агентство (Росавтодор)
Краткое описание источника данных	Источник данных представлен в виде таблицы с характеристиками по содержанию действующей сети автомобильных дорог общего пользования федерального значения (устройству защитных слоев, слоев износа и поверхностной обработки), их виду, со значением километра, описанием ремонтных работ, периодом их проведения, протяженностью участка и затратами на проведение ремонтных работ
Методология и изменения методологии для источника данных	ОДМ 218.3.082-2016 «Методические рекомендации по назначению технологий и периодичности проведения работ по устройству слоев износа и защитных слоев дорожных покрытий». ГОСТ Р 58862-2020 «Дороги автомобильные общего пользования. Ремонтные работы. Периодичность проведения»
Ссылки на методологию	<a href="https://docs.cntd.ru/document/456041198">https://docs.cntd.ru/document/456041198</a> <a href="https://docs.cntd.ru/document/1200173464">https://docs.cntd.ru/document/1200173464</a>

### Источник 2. Сведения об аварийно-опасных участках (место концентрации дорожно-транспортных происшествий) за отчетный год и планируемых мероприятиях по их ликвидации

Полное наименование источника данных	Сведения об аварийно-опасных участках (место концентрации дорожно-транспортных происшествий) за отчетный год и планируемых мероприятиях по их ликвидации
Сокращенное наименование источника данных	Форма 7а
Владелец источника данных	Федеральное дорожное агентство (Росавтодор)
Краткое описание источника данных	Данные являются массивом наблюдений, полученных из Формы 7а (Сведения об аварийно-опасных участках (место концентрации дорожно-транспортных происшествий) за отчетный год и планируемых мероприятиях по их ликвидации) за период с 2015 по 2020 г.
Методология и изменения методологии для источника данных	ОДМ 218.6.015-2015 «Рекомендации по учету и анализу дорожно-транспортных происшествий на автомобильных дорогах Российской Федерации»
Ссылки на методологию	<a href="https://docs.cntd.ru/document/1200120721">https://docs.cntd.ru/document/1200120721</a>