**MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA**

**Universitatea Tehnică a Moldovei**

**Facultatea Calculatoare, Informatică şi Microelectronică**

**Departamentul Informatică şi Ingineria Sistemelor**

Disciplina: AD

**Научная статья**

**Тема: “Анализ статистических данных электрических автомобилей”**

**A efectuat student (ul/a): \_\_\_\_\_\_\_\_\_\_\_\_**(I.Bahrin, gr. IA-214 )

**A controlat:\_\_\_\_\_\_\_\_\_\_\_\_\_\_**(asist.univ. V.Munteanu)

**Chişinău 2023**

**Оглавление**

[Абстракт 3](#_Toc153813389)

[Введение 3](#_Toc153813390)

[**1.** **Актуальность темы** 3](#_Toc153813391)

[**2.** **Цель** 3](#_Toc153813392)

[**3.** **Задание** 3](#_Toc153813393)

[Методы 4](#_Toc153813394)

[**4.** **Описание данных** 4](#_Toc153813395)

[**5.** **Описание модели** 5](#_Toc153813396)

[Результаты 6](#_Toc153813397)

[**6.** **Визуализация данных** 6](#_Toc153813398)

[**7.** **Визуализация выполнения модели** 10](#_Toc153813399)

[Выводы 11](#_Toc153813400)

[Библиография 12](#_Toc153813401)

# Абстракт

Это исследование анализирует статистическую информацию об электрических автомобилях, зарегистрированных в штате Вашингтон США. Анализ этого датасета поможет представить более полное представление о рынке автомобилей, потребностях потребителей и возможностях развития инфраструктуры для поддержки этих технологий.

Исследование основано на предположениях о зависимости различных факторов и фактических данных. Методология включает линейную регрессию для оценки влияния различных факторов, таких как марки, производители, год и другие категориальные переменные.

# Введение

## **Актуальность темы**

Анализ уровней популярности и распространенности электромобилей в штате Вашингтон остается значимым для понимания их динамики в различные периоды и среди разных возрастных групп. Это может быть важным для разработки программ и стратегий по расширению использования чистых автомобилей, улучшению экологической ситуации и поддержке инфраструктуры для электромобилей.

## **Цель**

Исследовать общие рыночные тенденции, потребительские предпочтения, оценка конкурентоспособности автомобилей.

## **Задание**

Изучение изменений в уровнях использования электромобилей в штате Вашингтон за последние десятилетия является важным для определения тенденций и различий в предпочтениях автомобильных технологий среди разных возрастных групп и в разные годы. Анализ этих данных может предоставить ценную информацию для разработки стратегий развития инфраструктуры зарядных станций, стимулирования экологически чистых решений в автомобильной индустрии и принятия решений в области экологической устойчивости.

# Методы

## **Описание данных**

Набор данных включает информацию об автомобилях, представленных в виде статистических регистрационных данных.

Данные охватывают период с 2008 по 2023 год и включают большое количество разнообразных автомобилей.

Набор данных содержит 150 000 записей.

Данные характеризуются несколькими категориальными переменными, включая марка автомобиля, производитель, тип двигателя и тд.

Переменные в рамках датасета об электромобилях имеют следующие значения:

1. **County:** Обозначает округ, в котором зарегистрирован автомобиль.
2. **City:** Город, в котором зарегистрирован автомобиль.
3. **State:** Штат, в котором зарегистрирован автомобиль.
4. **Model Year:** Год выпуска модели автомобиля.
5. **Make:** Производитель автомобиля.
6. **Model:** Модель автомобиля.
7. **Electric Vehicle Type:** Тип электромобиля (например, Battery Electric Vehicle - BEV, Plug-in Hybrid Electric Vehicle - PHEV и т.д.).
8. **Vehicle Location:** Местоположение автомобиля (например, координаты, точное место нахождения и т.д.).

Эти переменные предоставляют информацию о местоположении, характеристиках и типах электромобилей, что может быть полезным для анализа и планирования развития инфраструктуры для поддержки электромобилей в конкретных регионах или штатах.

## **Описание модели**

1. **Подготовка данных:** Вначале данные фильтруются и преобразуются для работы с категориальными переменными **Make**, **Model**, **Model Year**, и числовой переменной **Electric Range**. Они выбираются и приводятся к подходящему формату для анализа.
2. **Построение модели линейной регрессии:** **linear\_model <- lm(**Electric Range**~ Make + Model +**Model Year**, data = data\_filtered)**. Эта строка создает модель линейной регрессии, которая пытается объяснить изменения в **Electric Range** на основе выбранных переменных **Make**, **Model** и **Model Year**.
3. **Проверка сводной информации о модели:** **summary(linear\_model)**. Это выдает сводную статистику по построенной модели, включая коэффициенты регрессии, стандартные ошибки, значения t-статистики и p-значения для каждого коэффициента.
4. **Предсказания и оценка модели:** Затем создаются предсказанные значения **predictions <- data\_filtered %>% mutate(predicted = predict(linear\_model))** и оцениваются с использованием гистограммы ошибок предсказания, линейной диаграммы остатков и других графических представлений. Кроме того, вычисляются метрики оценки модели, такие как R-квадрат, среднеквадратическая ошибка (MSE), и средняя абсолютная ошибка (MAE).
5. **Графики зависимостей и распределения:** Код строит гистограммы, точечные и ящиковые диаграммы для визуализации связи между переменными и оценкой модели.

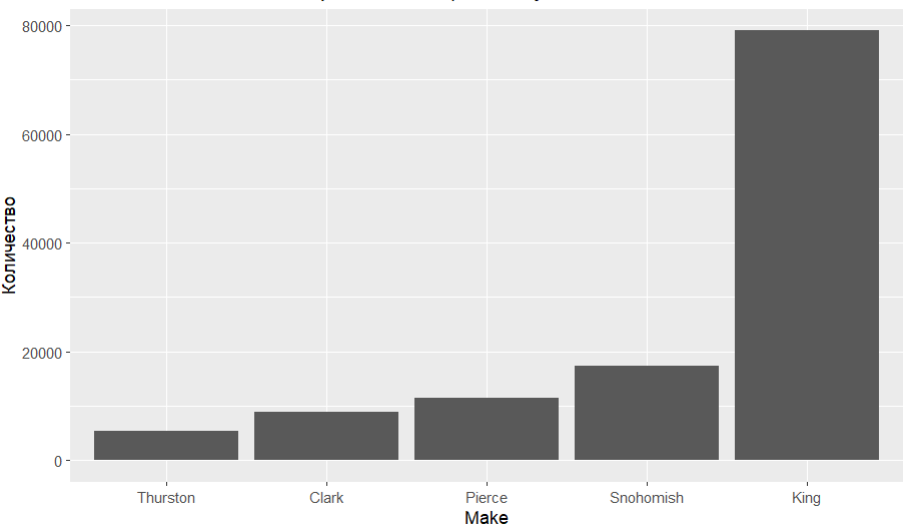
Эта модель линейной регрессии пытается определить, как переменные **Make**, **Model** и **Model Year** могут влиять на диапазон электрической подачи у электромобилей, а также оценить, насколько хорошо эта модель соответствует фактическим данным.

Переменные **Make**, **Model** и **Model Year** влияют на модель линейной регрессии как потенциальные предикторы зависимой переменной **Electric Range** (диапазон электрической подачи у электромобилей).

* **Make (Марка автомобиля):** Это категориальная переменная, представляющая производителя электромобиля. В модели эта переменная может влиять на **Electric Range** за счет того, что разные производители могут использовать различные технологии батарей или иметь разные дизайнерские особенности, которые влияют на эффективность батареи и, следовательно, диапазон езды.
* **Model (Модель автомобиля):** Эта категориальная переменная обозначает конкретную модель электромобиля. Разные модели могут иметь разные характеристики батарей, моторов и даже формы, что также может влиять на диапазон езды.
* **Model Year (Год модели):** Эта переменная представляет год выпуска модели автомобиля. С течением времени технологии в области батарей и моторов в электромобилях могут улучшаться, что может приводить к увеличению эффективности батареи и увеличению диапазона езды для более новых моделей.

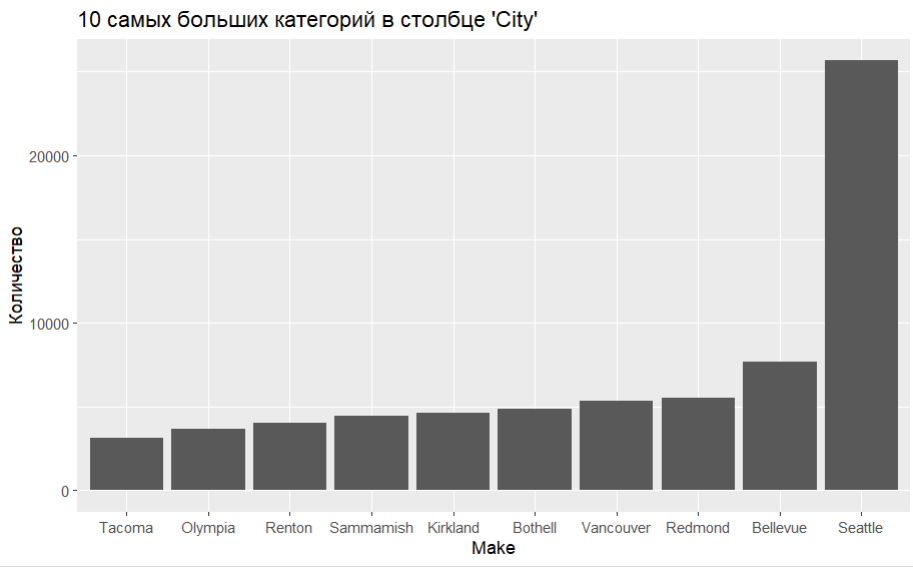
# Результаты

## **Визуализация данных**



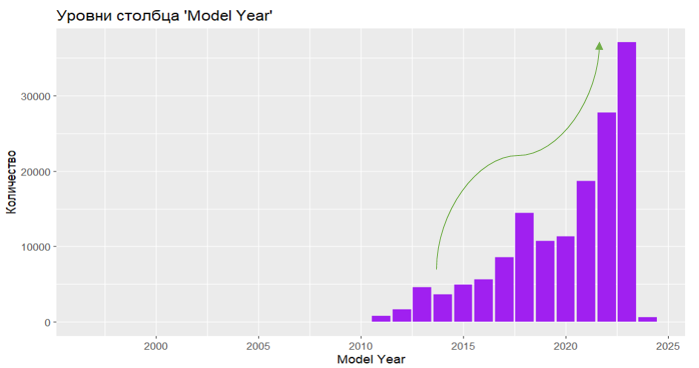
**Изображение 1** Топ 5 округов по кол-ву автомобилей

В округе Кинг, с населением около 80 000 электромобилей, наблюдается наибольшая популярность электрических автомобилей в штате Вашингтон. Это предоставляет потенциальные бизнес-возможности для развития зарядной инфраструктуры, продажи электромобилей и связанных с ними услуг.



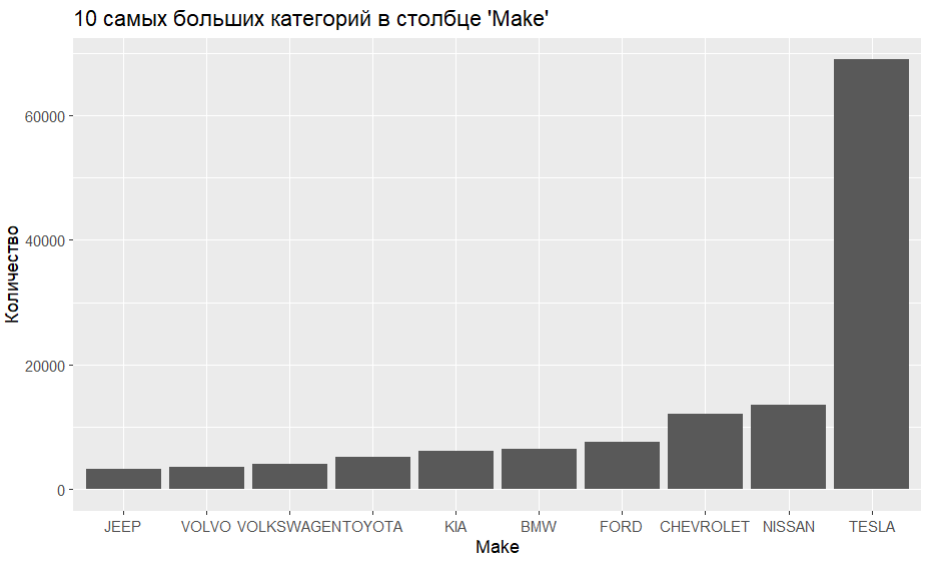
**Изображение 2** Топ 10 городов по кол-ву автомобилей

Знание о том, что в Сиэтле наибольшее количество электромобилей, позволяет автомобильным компаниям адаптировать свои стратегии к спросу на электромобили в этом регионе. Это помогает оптимизировать поставки, разрабатывать новые модели и строить зарядную инфраструктуру, улучшая конкурентоспособность и прибыльность бизнеса.



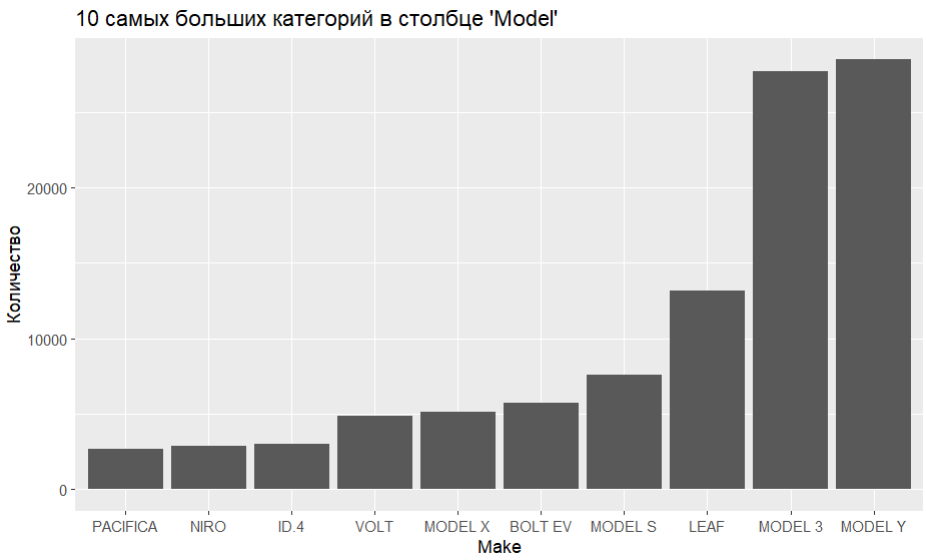
**Изображение 3** Года производства

Знание о том, что в 2023 году самые продаваемые электромобили в штате Вашингтон значительно увеличились по сравнению с предыдущими годами, предоставляет бизнесу ценную информацию. Это может способствовать реорганизации производства, а также акцентированию внимания на этой модели, что в итоге повысит прибыльность.



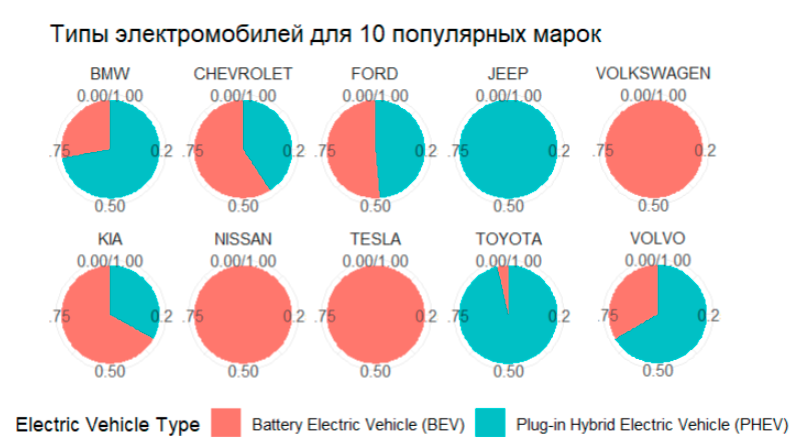
**Изображение 4** Топ 10 производителей автомобилей

График наглядно демонстрирует доминирование автомобилей Tesla на рынке. Это подчеркивает значительную популярность и влияние данного производителя в сегменте электромобилей, что может иметь важное значение для конкурентной стратегии в автомобильной индустрии.



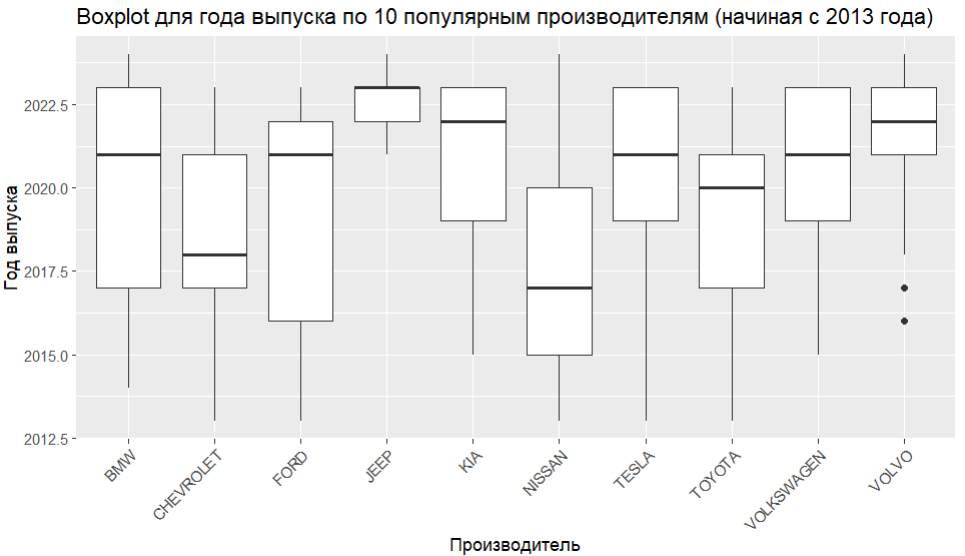
**Изображение 5** 10 самых популярных моделей

График 10 самых популярных марок автомобилей показывает, что Tesla Model 3 и Tesla Model Y сильно выделяются, занимая первые два места по популярности. Это может служить ключевой информацией для автомобильных производителей, стремящихся к конкурентоспособности на рынке электромобилей.



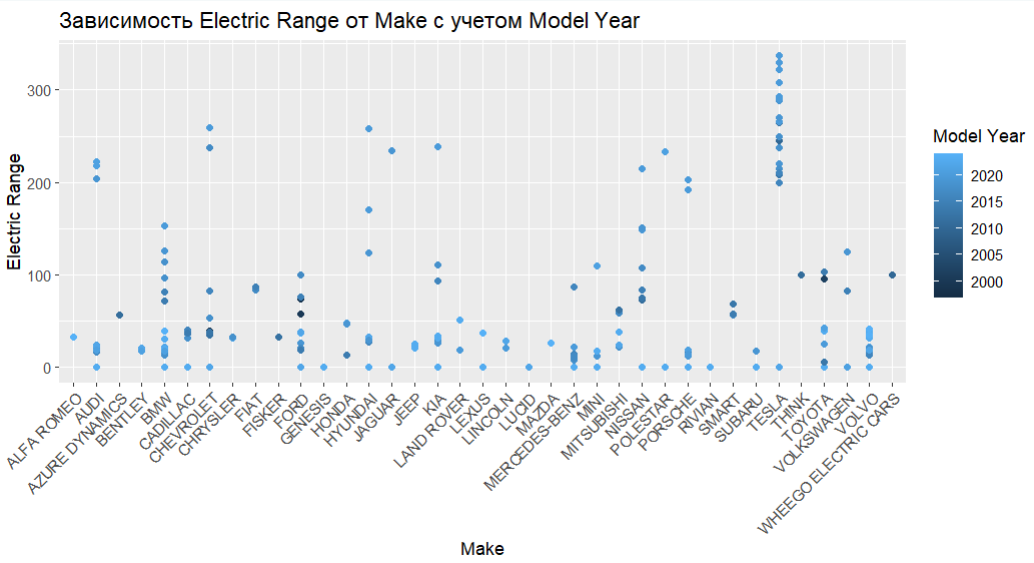
**Изображение 6** Распределение автомобилей по типу электродвигателя

График демонстрирует количество автомобилей производителей по типам электродвигателей в автомобилях может показать предпочтения между разными технологиями. Так же помогает увидеть статистику использования компаниями разных видов двигателей.



**Изображение 7** Количество выпущенных электрических авто

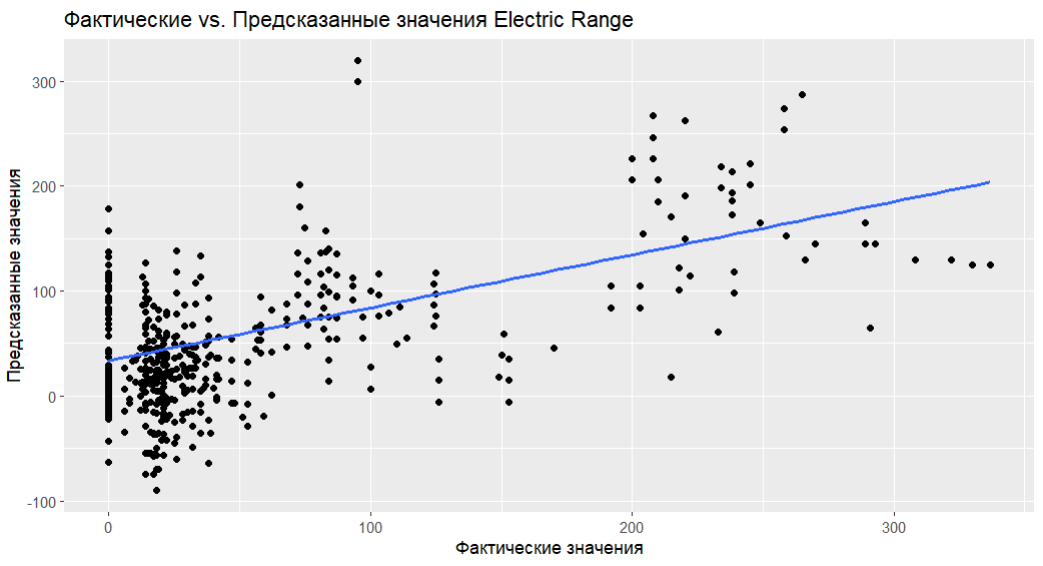
Boxplot для года выпуска автомобилей по 10 популярным производителям позволяет визуально сравнить распределение годов выпуска машин каждого производителя. Это помогает выявить тренды, диапазоны и выбросы в данных о производстве автомобилей.



**Изображение 8** График зависимости Electric Range от Make с учетом Model Year

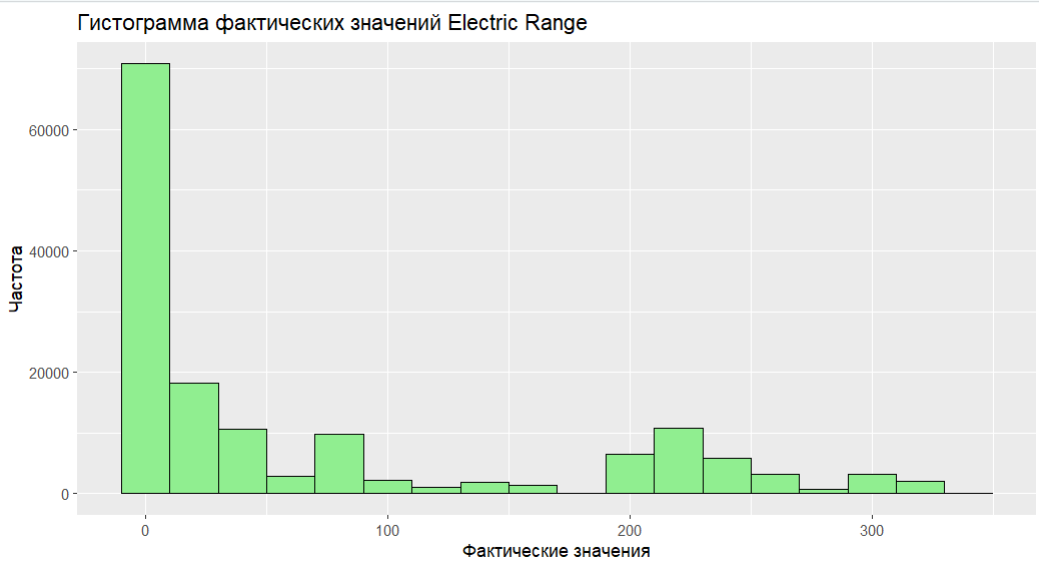
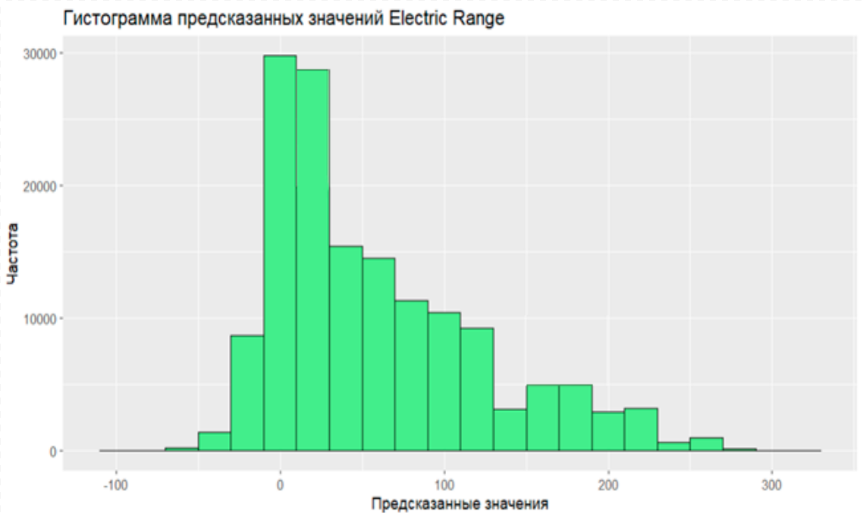
График зависимости Electric Range от Make с учетом Model Year иллюстрирует диапазон электрической ходимости (дальности пробега) автомобилей различных производителей в разные годы выпуска. Это помогает выявить тенденции в увеличении дальности пробега электромобилей с течением времени и между разными марками.

## **Визуализация выполнения модели**



**Изображение 9** Предсказание значений

Линейная регрессия для фактических и предсказанных значений Electric Range позволяет оценить точность модели. Сравнение этих значений демонстрирует, насколько модель точно предсказывает дальность пробега электромобилей на основе имеющихся данных, оценивая степень ее адекватности и точности.



**Изображение 10** **Гистограмма фактических/предсказанных значений**

# Выводы

**Бизнес и маркетинг:** Необходимо сфокусироваться на продажах и продвижении электромобилей с батареей большой емкости (BEV), так как они более популярны среди потребителей.

Развитие инфраструктуры зарядных станций в популярных городах, таких как Сиэтл, поможет привлекать больше клиентов.

**Для инженеров и автопроизводителей:** Необходимо инвестировать в исследования и разработку батарей с повышенной емкостью и уменьшенной стоимостью, чтобы сделать электромобили более доступными и увеличить дальность поездки.

Развитие инфраструктуры зарядных станций, особенно в популярных городах, поможет снизить барьеры для владельцев электромобилей.

**Для маркетинга:** Продвижение экологически чистых источников энергии, таких как солнечные батареи для зарядки электромобилей, может привлечь больше клиентов.

Специальные программы для владельцев электромобилей, такие как скидки на зарядку, могут стимулировать спрос.

**Общий вывод:** Рынок электромобилей в Вашингтоне растет, особенно в городе Сиэтл.

Успех на растущем рынке потребует фокусировки на BEV, инновации в технологии и инфраструктуре зарядных станций.

# Библиография

1. <https://catalog.data.gov/dataset/electric-vehicle-population-data> accesat la 27.11.2023
2. <https://www.projectpro.io/article/top-10-machine-learning-projects-for-beginners-in-2021/397> , accesat la 28.11.2023
3. <https://towardsdatascience.com/14-kaggle-competitions-to-start-your-data-science-journey-41943496b6f4> , accesat la 29.11.2023