**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 ЦЕН АВТОМОБИЛЕЙ ОТНОСИТЕЛЬНО ХАРАКТЕРИСТИК”**

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

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

**Chişinău 2023**

**Оглавление**

[**Абстракт** 3](#_30j0zll)

[**Введение** 3](#_1fob9te)

[1. Актуальность темы 3](#_3znysh7)

[2. Цель 3](#_2et92p0)

[3. Задание 3](#_tyjcwt)

[**Методы** 4](#_3dy6vkm)

[4. Описание данных 4](#_1t3h5sf)

[5. Описание модели 4](#_4d34og8)

[**Результаты** 6](#_2s8eyo1)

[6. Визуализация данных 6](#_17dp8vu)

[7. Визуализация выполнения модели](#_3rdcrjn) 9

[**Выводы**](#_26in1rg) 10

[**Библиография** 1](#_lnxbz9)1

# Абстракт

Это исследование анализирует цену автомобилей относительно его характеристик на рынке Индии. Целью работы является выявление цены от его характеристик и на сколько сильно это будет отличаться от реальной цены на автомобиль.

Исследование основано на предположениях о зависимости цены от характеристик автомобиля и его состоянии. Методология включает линейную регрессию для оценки влияния различных факторов, таких как пробег, год и другие категориальные переменные.

# Введение

## **Актуальность темы**

Анализ цен автомобилей в зависимости от их характеристик актуален из-за быстрого развития автомобильного рынка, необходимости информированного выбора для покупателей, оптимизации ценообразования для продавцов и прогресса в аналитических технологиях, подчеркивая его значимость в контексте экономических и социальных изменений, связанных с автомобильной индустрией.

## **Цель**

Исследовать общие тенденции в ценообразование автомобилей на основе их характеристик для прогнозирования цены автомобиля.

## **Задание**

Необходимо проанализировать цены и характеристики автомобилей, чтобы понять тенденции и вариации в их ценообразовании. Этот анализ может предоставить ценные данные для платформ по продаже автомобилей.

# Методы

## **Описание данных**

Набор данных включает информацию о ценах и характеристиках автомобилей на рынке Индии.

Дополнительные столбцы данных включают в себя индикатор, единицы измерения, год и флаги для оценки качества данных.

Набор данных содержит 6 014 записей.

Данные характеризуются несколькими категориальными и числовыми переменными.

Качество данных отмечается флагами, и есть некоторые отсутствующие значения.

* Year: Год выпуска автомобиля..
* Kilometers Driven: Пробег автомобиля в километрах.
* Fuel Type: Тип топлива, на котором работает автомобиль.
* Transmission: Тип коробки.
* Owner Type: Сведения о количестве владельцев.
* Mileage: Расход топлива автомобиля.
* Engine: Объем двигателя в кубических сантиметрах (см³).
* Power: Мощность двигателя в лошадиных силах (л.с.).
* Seats: Количество сидений в автомобиле.

## **Описание модели**

Эта модель, основанная на линейной регрессии, стремится объяснить изменения цены в зависимости от нескольких факторов, таких как пробег, годя, мощность автомобиля и другие категориальные переменные.

Переменные и их влияние на цену:

* Пробег автомобиля:

-Kilometers Driven :

- Пробег каждого автомобиля оказывает различное влияние на его ценность. Некоторые автомобили со значительным пробегом не показывают статистически значимого воздействия на цену (p-value > 0.05), однако общая тенденция свидетельствует о том, что стоимость автомобиля уменьшается по мере увеличения его пробега..

* Год:

- Year:

- Год выпуска автомобиля значимо влияет на его цену, при этом коэффициент статистически значим при текущих уровнях достоверности (p-value > 0.1), что подтверждает важность года выпуска как фактора, оказывающего существенное влияние на рыночную стоимость автомобиля.

* Расход топлива:

- Mileage:

- Расход топлива каждого автомобиля оказывает свое влияние на его стоимость. Автомобили с высоким расходом топлива часто имеют более низкую рыночную цену, хотя этот эффект может не всегда быть статистически значимым (p-value > 0.05). Однако общая тенденция указывает на то, что автомобили с более эффективным расходом топлива, как правило, ценятся выше на рынке.

Другие категориальные переменные:

- Seats, Owner Type, Transmission и Fuel Type :

- Категории этих переменных не значимо влияют на цену автомобиля.

Взаимодействия:

- Есть взаимодействия между Year и Kilometers Driven, что указывает на взаимосвязь между этими переменными и ценной.

Качество модели:

- Multiple R-squared: 0.6899

- Это показывает, насколько хорошо модель соответствует данным. В данном случае, модель объясняет около 68.99% изменчивости цены от характеристик.

- F-статистика: 704.7, \*p\*-value: < 2.2e-16

- Этот тест оценивает значимость всей модели. Низкое значение \*p\*-value (< 0.05) указывает на то, что по крайней мере, один из предикторов (переменных) вносит значимый вклад в объяснение изменений уровня смертности от самоубийств.

Модель имеет значимые переменные, но следует учитывать, что она объясняет менее половины изменчивости данных, поэтому другие факторы могут также играть важную роль в преобразовании цены.

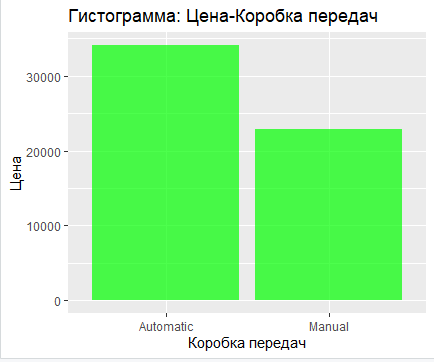
# Результаты

## **Визуализация данных**

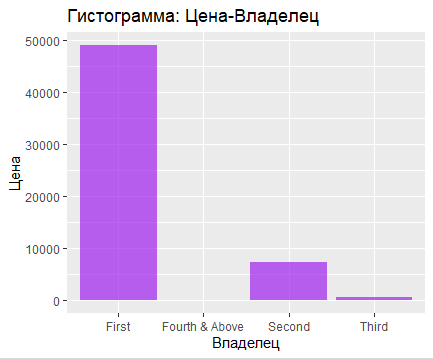


**Изображение 1** Визуальный анализ столбца Year.

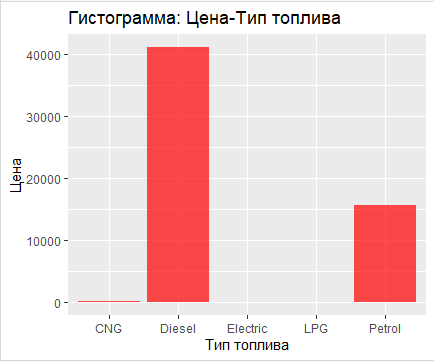
Данный график отображает взаимосвязь между ценой и годом выпуска автомобиля. График указывает на то, что чем новее автомобиль тем он дороже.



**Изображение 2** Визуальный анализ столбца Transmission



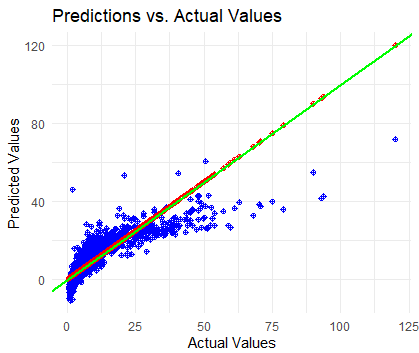
**Изображение 3** Визуальный анализ столбца Owner Type



**Изображение 3** Визуальный анализ столбца Fuel Type

Данные примеры графиков отражают значимость категориальных данных на цену автомобиля, но стоить учесть, что эти данные часто зависят от рынка сбыта, так как у каждого региона свои потребности и ожидание от автомобиля. В данном случае мы наблюдаем значимость для рынка Индии.

## **Визуализация выполнения модели**



**Изображение 4** Предсказание значений

Данный график отображает предсказанные цены. Глядя на него можно сделать вывод, что из-за слабой модели, предсказанные данные отличаются от действительных, но можно увидеть закономерность того, что предсказанные данные варьируются в определённом промежутке и не уходят слишком далеко от действительных данных.

# Выводы

Исследование подчеркивает, что характеристики автомобиля, такие как год выпуска, тип топлива, коробка передач, пробег и количество владельцев, играют ключевую роль в определении его рыночной стоимости, подчеркивая сложность процесса ценообразования на автомобильном рынке.

Влияние факторов, таких как год выпуска, тип топлива, коробка передач, пробег и количество владельцев, подтверждает сложность процесса ценообразования.

В дальнейших исследованиях предлагается расширить анализ, включив более детальные аспекты, такие как влияние бренда и модели автомобиля, а также региональные особенности рынка. Это позволит получить более глубокое понимание механизмов формирования цен и их вариаций в зависимости от различных рыночных условий и потребительских предпочтений, что важно как для продавцов, так и для покупателей в процессе принятия обоснованных решений о купле-продаже автомобилей.

# Библиография

1. <https://www.kaggle.com/> ,

accesat la 18.12.2023

1. https://www.projectpro.io/article/top-10-machine-learning-projects-for-beginners-in-2021/397 , accesat la 18.12.2023
2. <https://towardsdatascience.com/14-kaggle-competitions-to-start-your-data-science-journey-41943496b6f4> , accesat la 18.12.2023