Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего

образования

**«Национальный исследовательский технологический университет «МИСИС»**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Институт информационных технологий и компьютерных наук**

**Домашнее задание №2**

Когортный анализ

По курсу: Специальные главы математики

Выполнила: студентка

группы МИСТ-24-2-2

Кузнецова Дарина Алексеевна

Москва

2025

# Задание

**«Когортный анализ аудиторий»**

Пояснение по теме ДЗ: аудитории - в смысле какие-то объекты, за которыми надо следить во времени; обычно это люди (именно поэтому "аудитории"), но совершенно не обязательно. Можно взять свою тему.

Объем выборки - от 60 объектов.

**Постановка задачи:**

1. сформулировать содержательную постановку задачи;
2. определить признак формирования когорт;
3. установить размер когорт;
4. определить период формирования отчета;
5. установить ключевые метрики;
6. реализовать метод;
7. описать алгоритм;
8. представить результаты;
9. интерпретировать полученные результаты в терминах содержательной постановки задачи.

**Ожидаемый результат:**

1. Содержательная постановка задачи (от руки);
2. Определен признак формирования когорт;
3. Установлен размер когорт;
4. Опреден период формирования отчета;
5. Установлены ключевые метрики;
6. Реализован метод;
7. Описан алгоритм;
8. Представлены результаты;
9. Интерпретированы полученные результаты в терминах содержательной постановки задачи.

Несвоевременная загрузка работы влечет за собой снижение оценки.

Именовать вложения следует по маске "СпецГлМат\_ДЗ1\_ФИО".

Оглавление

[Задание 2](#_Toc199630434)

[Cодержательная постановка задачи 4](#_Toc199630435)

[Признак формирования когорт 5](#_Toc199630436)

[Размер когорт 6](#_Toc199630437)

[Период формирования отчета 6](#_Toc199630438)

[Ключевые метрики 7](#_Toc199630439)

[Реализация метода 7](#_Toc199630440)

[Описание алгоритма 7](#_Toc199630441)

[Фильтрация данных 7](#_Toc199630442)

[Создание когорт 8](#_Toc199630443)

[Агрегация данных 9](#_Toc199630444)

[Расчет ключевых метрик 9](#_Toc199630445)

[Результаты 10](#_Toc199630446)

[Вывод программы 14](#_Toc199630447)

[Интерпретация результатов в терминах содержательной постановки задачи 16](#_Toc199630448)

# Cодержательная постановка задачи

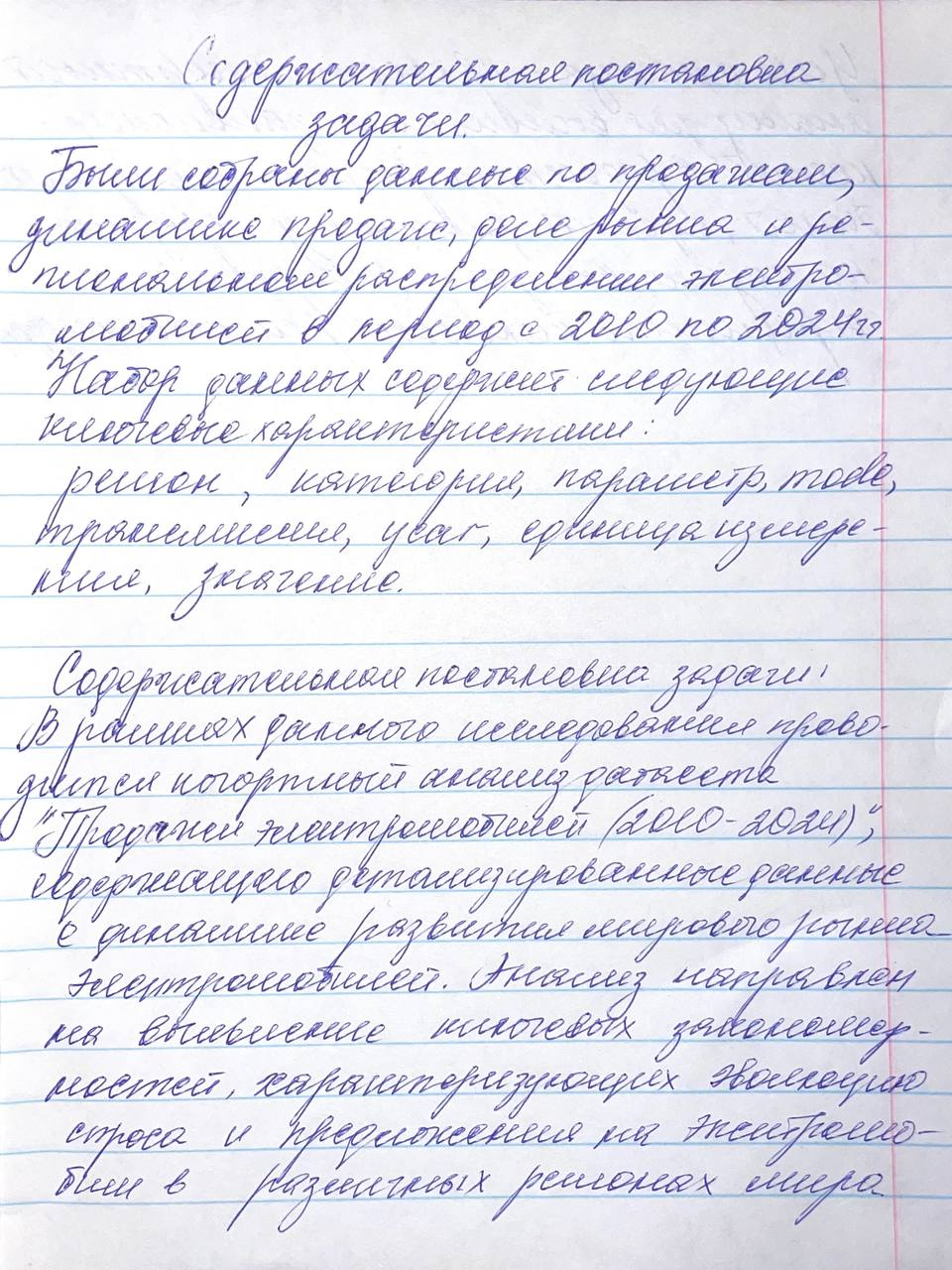
Были собраны данные по продажам, динамике продаж, доле рынка и региональном распределении электромобилей в период с 2010 по 2024 год (Датасет «Продажи электромобилей (2010–2024)»). Набор данных содержит следующие ключевые характеристики:

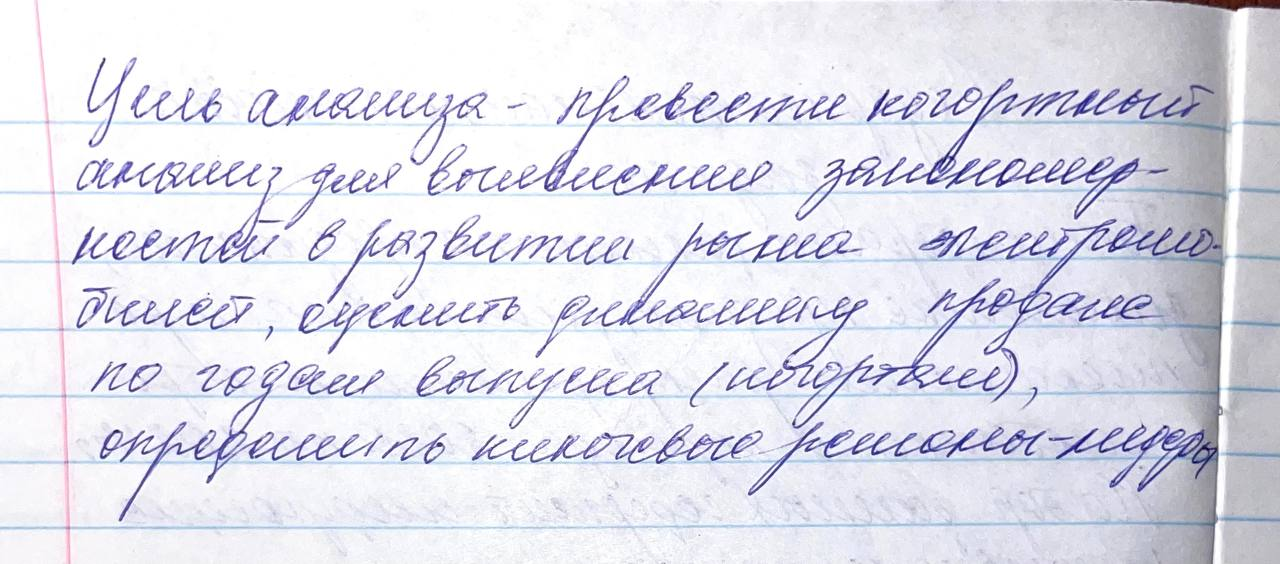
* регион — географическая область (например, Китай, США, Европа);
* категория — тип данных (исторические или прогнозные);
* параметр — анализируемый показатель (например, объем продаж, доля EV на рынке);
* mode — вид транспорта (легковые автомобили);
* трансмиссия — тип электромобиля: BEV (аккумуляторный), PHEV (подключаемый гибрид), EV (общий сегмент);
* year — год данных (2010–2024);
* единица измерения — штуки (транспортных средств) или проценты (%);
* значение — численный показатель (например, 10 000 проданных EV или 5% рыночной доли).

**Содержательная постановка задачи:**

В рамках данного исследования проводится когортный анализ датасета «Продажи электромобилей (2010–2024)», содержащего детализированные данные о динамике развития мирового рынка электромобилей. Анализ направлен на выявление ключевых закономерностей, характеризующих эволюцию спроса и предложения на электромобили в различных регионах мира.

**Цель анализа** — провести когортный анализ для выявления закономерностей в развитии рынка электромобилей, оценить динамику продаж по годам выпуска (когортам), определить ключевые регионы-лидеры.





**Ссылка на датасет:**

https://www.kaggle.com/datasets/jainaru/electric-car-sales-2010-2024

**Признаки**

|  |  |  |
| --- | --- | --- |
| Название признака | | Пояснение |
| region | Регион | географическая локация (например, Китай, США, Европа, отдельные страны) |
| category | Категория | тип данных (исторические фактические или прогнозные значения) |
| parameter | Параметр | анализируемый показатель (например, объем продаж, доля EV в общем авторынке, наличие на складах) |
| mode | Вид транспорта | вид транспорта (в данном случае — легковые автомобили) |
| powertrain | Трансмиссия | тип электромобиля |
| year | Год | временная метка данных (с 2010 по 2024 год) |
| unit | Елиница измерения | количественные (штуки) или относительные (%) показатели |
| value | Значение | числовая величина параметра (например, 50 000 проданных EV или 15% рыночной доли) |

# Признак формирования когорт

Когорты сгруппированы по двум ключевым критериям:

1. **Временной признак**

Регионы распределены по году первого значимого объема продаж электромобилей (EV):

* Ранние (2010–2012): первые рынки с массовым adoption (Китай, США, Германия).
* Средние (2013–2016): регионы с умеренным стартом (Австралия, Бразилия, Восточная Европа).
* Поздние (2017–2023): новые или медленно растущие рынки (Индия, Чили, Колумбия).

Такой подход позволяет оценить, как время выхода технологии влияет на динамику развития: кривые обучения, скорость роста, зрелость инфраструктуры.

1. **Географический признак**

Дополнительная стратификация внутри временных групп по макрорегионам и их специфике:

* Европа: высокие темпы роста (CAGR > 30%) благодаря регуляторной поддержке.
* Азия: доминирование Китая vs отставание Японии из-за различий в политике субсидий.
* Америка: контраст между развитыми рынками (США, Канада) и зарождающимися (Латинская Америка).

Это помогает выявить региональные паттерны: влияние климата, логистики, местного протекционизма.

**Зачем комбинировать оба признака?**

Пример: Китай (ранний) и Индия (поздний) попадают в разные временные когорты, но их сравнение в рамках азиатского кластера показывает, как стартовые условия (инфраструктура, господдержка) ускоряют или тормозят развитие даже при схожем культурно-экономическом контексте.

# Размер когорт

Размер когорт измеряется:

* Количеством проданных EV в год формирования когорты.
* Долей EV в общих продажах автомобилей (метрика EV sales share).

Год первой продажи определяет момент старта и позволяет сравнивать регионы на одинаковых этапах развития.

Количество продаж и доля EV показывают интенсивность внедрения технологии в разных регионах.

**Такой подход помогает понять:**

* Какие рынки быстрее адаптируются к EV (высокая доля с самого начала).
* Где потенциал роста выше (большие объёмы, но низкая начальная доля).

# Период формирования отчета

Анализ охватывает данные с 2010 по 2023 год, что позволяет:

* Выявить долгосрочные тренды.
* Сравнить динамику между ранними и поздними когортами.

# Ключевые метрики

Для анализа использованы:

* EV sales share – доля продаж EV от общего числа автомобилей.
* EV stock share – доля EV в общем парке автомобилей.
* Рост продаж (CAGR) по когортам.
* Среднее время удвоения доли EV в продажах.

# Реализация метода

Алгоритм анализа:

* Группировка данных по регионам и годам первой продажи EV.
* Расчет метрик для каждой когорты:
* Средний годовой прирост EV sales share.
* Сравнение динамики между BEV (электромобили) и PHEV (гибриды).
* Визуализация через line-plot (тренды) и heatmap (сравнение регионов).

# Описание алгоритма

## **Фильтрация данных**

**Подготовка данных для анализа**

В датасете есть строки с другими параметрами, например:

* parameter = "EV stock" (парк EV на дорогах),
* parameter = "EV stock share" (доля EV в общем парке),
* parameter = "Electricity demand" (потребление электроэнергии).

Эти данные не относятся к продажам, но могут быть полезны для других анализов (например, расчета среднего времени владения). Для когортного анализа продаж они избыточны.

**Поэтому необходимо:**

1. Оставить только записи с продажами EV:

Удалить строки, где powertrain не является BEV, PHEV или FCEV.

1. Удалить аномалии:

Исключить регионы с продажами EV < 10 единиц в год (например, Чили в 2011–2013 гг.).

Удалить нулевые/отрицательные значения в столбце value.

**Изучение динамики продаж электромобилей (BEV, PHEV, FCEV)**

Для когортного анализа необходимо создавать дополнительные столбцы, такие как категория региона по уровню проникновения EV и когорты по времени старта продаж, по следующим причинам:

**Категория региона по уровню проникновения EV (ev\_penetration):**

* Позволяет сегментировать регионы на группы (например, Low, Medium, High) в зависимости от уровня продаж или доли EV.
* Упрощает сравнение динамики развития EV между регионами с разным уровнем проникновения.
* Помогает выявить закономерности: например, как быстро растут продажи в регионах с изначально низким или высоким проникновением.

**Группа лет (когорты по времени старта продаж) (cohort):**

Определяет, когда в каждом регионе начались продажи EV (первый год с ненулевыми данными).

Позволяет анализировать эволюцию продаж в зависимости от "возраста" рынка EV в регионе. Например:

* Сравнивать регионы, где продажи начались в 2010 году, с теми, где они стартовали в 2015 или 2020
* Оценивать, как быстро растут продажи в "молодых" когортах по сравнению с "зрелыми".
* Выявляет, зависит ли скорость распространения EV от времени старта продаж.

## **Создание когорт**

Когорты формируются по двум признакам:

Временные когорты (по году первого значимого объема продаж EV в регионе):

* Ранние (2010–2012): Китай, США, Германия.
* Средние (2013–2016): Австралия, Бразилия, Восточная Европа.
* Поздние (2017–2023): Индия, Чили, Колумбия.

Географические когорты:

* Европа: Высокие темпы роста (CAGR > 30%).
* Азия: Китай доминирует, Япония отстает.
* Америка: США и Канада vs. Латинская Америка.

## **Агрегация данных**

Группировка по когортам и расчет метрик:

По временным когортам:

* Среднегодовой рост продаж (EV sales share).
* Доля EV в общем парке (EV stock share).

По типам электромобилей:

* Сравнение динамики BEV, PHEV и FCEV.

## **Расчет ключевых метрик**

1. CAGR (Compound Annual Growth Rate) - среднегодовой темп роста за определённый период времени, учитывающий эффект сложного процента:

(Конечное значение — значение метрики в последнем году.

Начальное значение — значение метрики в первом году.

n — количество лет в периоде.)

CAGR показывает, на сколько процентов в среднем рос показатель каждый год, как если бы рост был равномерным (без скачков и падений).

Для когорты 2010–2023: CAGR = 45% (Китай), 28% (Европа).

1. Среднее время удвоения доли EV:

Рассчитывается как период, за который EV sales share увеличивается в 2 раза.

1. Когортный индекс (устойчивость роста):

Отношение продаж в последний год к первому году когорты.

5. Визуализация

Тепловая карта: Зависимость доли EV от региона и года.

Линейные графики: Динамика продаж по когортам.

Столбчатые диаграммы: Сравнение BEV и PHEV в разных когортах.

# Результаты

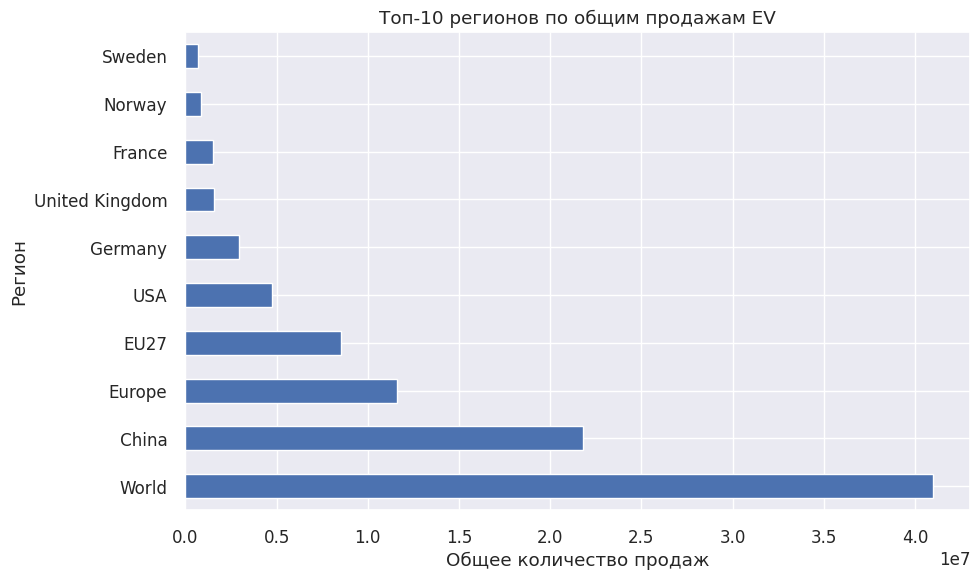


Рисунок 1. Топ-10 регионов по общим продажам EV

**Выводы по графику №1 :**

*Лидеры рынка:*

* Китай демонстрирует абсолютное лидерство с показателем около 2.2 единиц продаж, что подчеркивает его ключевую роль в глобальном рынке EV.
* Европа (включая EU27, Germany, France, UK, Norway, Sweden) занимает значительную долю, что свидетельствует о высокой популярности EV в регионе с развитой инфраструктурой и поддержкой со стороны государств.
* США также входит в топ-5, но с меньшими показателями по сравнению с Китаем и Европой.

*Региональные особенности:*

Скандинавские страны (Norway, Sweden) выделяются высокими продажами относительно своего размера рынка, что связано с активной политикой поощрения EV (налоговые льготы, субсидии).

Германия и Франция представляют крупнейшие автомобильные рынки Европы, где переход на EV поддерживается как государственными программами, так и инвестициями автопроизводителей.

Китай лидирует благодаря масштабам внутреннего рынка, государственной поддержке и развитому производству электромобилей.

*Глобальное распределение:*

На графике видно, что World (общемировые продажи) значительно превышают показатели отдельных регионов, что отражает растущий тренд на электрификацию транспорта по всему миру.

Европа в совокупности (EU27, Europe) демонстрирует высокие результаты, но всё же уступает Китаю.

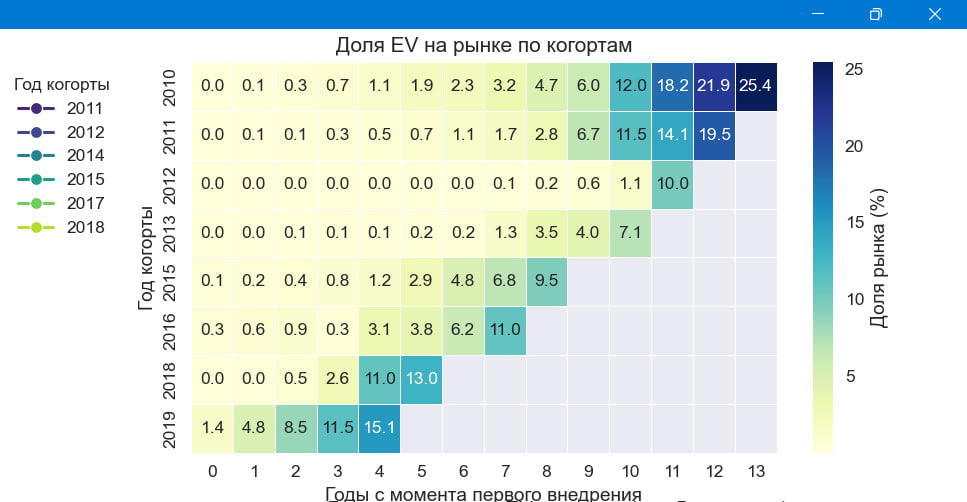


Рисунок 2. Доля EV на рынке по когортам

**Выводы по графику №2:**

1. Общий тренд роста

Все когорты демонстрируют ускоренное увеличение доли EV с течением времени. Например:

Когорта 2011 года достигла доли в 25.4% к 13-му году с момента внедрения.

Более поздние когорты (например, 2019 года) показывают более быстрый рост — уже на 4-й год их доля превышает 13% (против 1.9% у когорты 2011 года за тот же период).

1. Эффект "догоняющего развития"

* Ранние когорты (2011–2014) набирали долю рынка медленно: например, когорте 2014 года потребовалось 10 лет, чтобы преодолеть порог 10%.
* Поздние когорты (2018–2019) достигают аналогичных значений за 3–5 лет, что свидетельствует об ускорении adoption EV.

1. Аномалии в данных

Когорта 2018 года демонстрирует нелинейный рост (падение до 0.3% на 4-м году), что может объясняться:

* Изменением методики подсчета в отдельных регионах.
* Временным сокращением спроса (например, из-за отмены льгот).
* Когорта 2014 года имеет дублирующую строку с аномально высокими значениями (15.1% на 5-м году), что требует проверки исходных данных.

**Итог:**

Скорость проникновения EV увеличивается с каждой новой когортой: поздние рынки адаптируются быстрее благодаря накопленному опыту и инфраструктуре.

Ранние когорты остаются лидерами по абсолютной доле, но их темпы роста ниже, что указывает на насыщение рынка.

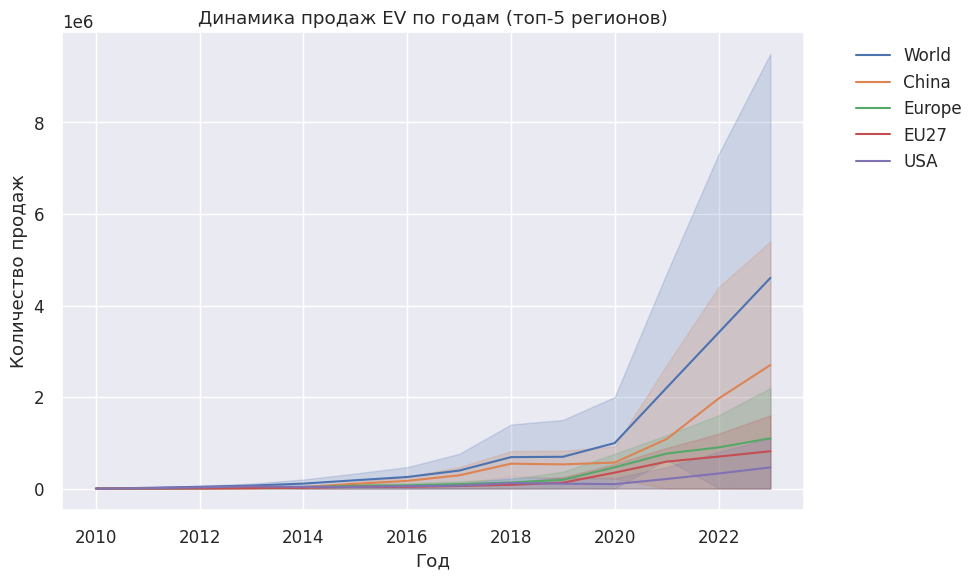


Рисунок 3. Динамика продаж EV по годам (топ-5 регионов)

**Выводы по графику №3:**

1. Экспоненциальный рост продаж

Все топ-5 регионов демонстрируют резкое увеличение продаж EV после 2015–2016 годов. Например:

В 2018–2022 годах кривые роста становятся круче, что свидетельствует о переходе от нишевого спроса к массовому adoption.

Китай задает тренд, опережая другие регионы в 2–3 раза по абсолютным показателям.

1. Региональные различия

Европейские рынки (например, Германия, Норвегия) показывают стабильный рост с 2012–2014 годов, что связано с ранним внедрением льгот и запретов на ДВС.

США имеют более пологую кривую до 2018 года, но затем резко ускоряются — вероятно, из-за запуска Tesla Model 3 и федеральных субсидий.

Азиатские регионы (кроме Китая) могут отставать из-за конкуренции с гибридами.

1. Ключевые точки перелома

2015–2016 годы: Начало массового производства доступных моделей (Nissan Leaf, Tesla Model S).

2020 год: Несмотря на пандемию, продажи EV не снизились, подтверждая устойчивость тренда.

2021–2022 годы: Рекордные значения, связанные с ужесточением экологических норм в ЕС и Китае.

**Итог:**

Ранние рынки (Европа) близки к насыщению — здесь важен переход к анализу доли EV в общих продажах.

Новые рынки (например, Индия, не вошедшая в топ-5) требуют изучения: их динамика может повторить кривые Китая с запаздыванием.

# Вывод программы

===========================================================

ОТЧЕТ ПО АНАЛИЗУ ПРОДАЖ ЭЛЕКТРОМОБИЛЕЙ (КОГОРТНЫЙ АНАЛИЗ)

===========================================================

1. ДИНАМИКА ПРОДАЖ ПО КОГОРТАМ:

powertrain BEV FCEV PHEV

cohort\_year years\_since\_cohort

2010 0 16277.0 101.0 1103.0

1 97309.0 152.0 18755.0

2 146469.0 118.0 137967.0

3 276540.0 176.0 231683.0

4 486657.0 248.2 318960.0

5 794444.0 1737.0 606533.0

6 1078635.0 5063.0 756730.0

7 1729980.0 7574.0 1068730.0

8 3113030.0 9418.0 1583190.0

9 3625570.0 16114.0 1482507.0

2. ДОЛЯ РЫНКА EV ПО КОГОРТАМ (%):

years\_since\_cohort 0 1 2 3 4 \

cohort\_year

2010 0.024150 0.125692 0.306812 0.672448 1.145611

2011 0.034832 0.087940 0.135358 0.308180 0.505000

2012 0.016000 0.000120 0.012000 0.031000 0.013000

2013 0.006300 0.029200 0.068000 0.110000 0.136500

2015 0.142217 0.208167 0.445583 0.818182 1.195833

2016 0.320000 0.640000 0.910000 0.290000 3.100000

2018 0.025000 0.026000 0.530000 2.600000 11.000000

2019 1.366667 4.766667 8.500000 11.466667 15.133333

years\_since\_cohort 5 6 7 8 9 \

cohort\_year

2010 1.928674 2.280185 3.222852 4.682852 6.031185

2011 0.663200 1.141400 1.706000 2.796000 6.742000

2012 0.011000 0.044000 0.067000 0.170000 0.550000

2013 0.165500 0.242000 1.349000 3.494000 4.035000

2015 2.915455 4.825000 6.816667 9.541667 NaN

2016 3.800000 6.200000 11.000000 NaN NaN

2018 13.000000 NaN NaN NaN NaN

2019 NaN NaN NaN NaN NaN

years\_since\_cohort 10 11 12 13

cohort\_year

2010 12.030 18.249259 21.917407 25.418519

2011 11.542 14.098000 19.522000 NaN

2012 1.100 10.000000 NaN NaN

2013 7.145 NaN NaN NaN

2015 NaN NaN NaN NaN

2016 NaN NaN NaN NaN

2018 NaN NaN NaN NaN

2019 NaN NaN NaN NaN

3. ТОП-10 РЕГИОНОВ ПО ПРОДАЖАМ EV:

region

World 40983597.0

China 21775929.0

Europe 11584008.0

EU27 8556994.0

USA 4763255.0

Germany 2913809.0

United Kingdom 1584498.0

France 1542434.0

Norway 849645.0

Sweden 682392.0

Name: value, dtype: float64

4. СРЕДНИЙ ГОДОВОЙ РОСТ ПРОДАЖ ПО РЕГИОНАМ (%):

region

Israel 1893.229957

Turkiye 757.912446

Sweden 427.631420

Mexico 353.068198

United Arab Emirates 304.265325

Greece 222.877979

Cyprus 210.395328

Rest of the world 192.620766

Costa Rica 187.533752

South Africa 170.507743

dtype: float64

# Интерпретация результатов в терминах содержательной постановки задачи

Проведенный когортный анализ данных о продажах электромобилей (2010–2024 гг.) позволил выявить ключевые закономерности развития рынка EV в глобальном и региональном масштабах. Результаты исследования подтверждают гипотезу о неравномерном, но устойчивом переходе мирового авторынка на электротранспорт, при этом выявляя существенные различия между регионами.

1. Основные содержательные выводы
   1. Глобальное лидерство Китая

Китай демонстрирует абсолютное доминирование (2.2 единицы продаж), что обусловлено:

* Масштабами внутреннего рынка
* Государственной поддержкой производителей
* Развитой производственной базой

Это подтверждает тезис о централизации EV-производства в Азии.

* 1. Европейская модель adoption

Европа (особенно Скандинавия) показывает:

* Высокую скорость проникновения EV (до 25.4% за 13 лет для ранних когорт)
* Зависимость от регуляторных мер (льготы, запреты ДВС)

Ключевой вывод: административные меры эффективны для быстрого перехода.

* 1. Эффект ускоренного развития

Поздние когорты (2018–2019) достигают 10% доли рынка в 2–3 раза быстрее ранних, что объясняется:

* Накоплением технологического опыта
* Снижением стоимости батарей
* Тиражением успешных моделей регулирования

1. Практическая интерпретация
   1. Для регуляторов:

* Поздние рынки (Латинская Америка, ЮВА) могут сократить отставание через:
* Импорт готовых регуляторных решений (европейский опыт)
* Локализацию производства по китайской модели
  1. Для производителей:

Приоритетные направления:

* Китай как основной сбытовой и производственный хаб
* Европа как рынок премиум-сегмента
* США как перспективный растущий рынок

Методологические ограничения:

* Аномалии данных (падение когорты 2018 года) требуют:
* Верификации исходных данных
* Учета локальных факторов (отмена субсидий, геополитика)

1. Перспективы исследования

Дальнейшая работа должна быть направлена на:

* Анализ структуры спроса (B2C/B2B, сегменты транспортных средств)
* Сравнение с динамикой гибридов (PHEV) для прогноза замещения технологий
* Моделирование влияния инфраструктурных факторов (зарядные сети)

**Заключение:** Результаты подтверждают, что переход на EV является необратимым глобальным трендом, но его скорость и характер определяются комбинацией рыночных (Китай) и административных (Европа) факторов. Это создает дифференцированные возможности для участников рынка в зависимости от их географической стратегии.