

Анализ частоты и интенсивности погодных явлений, ведущих к экономическим потерям в различных регионах РФ

1 Введение

1.1 Цель проекта

Данный проект направлен на анализ данных о погодных явлениях в России с целью выявления их влияния на экономические потери в различных федеральных округах. В проекте используются данные о погоде, которые позволяют оценить частоту и интенсивность различных явлений, а также выявить регионы и временные периоды, подверженные наибольшим рискам экономических убытков.

1.2 Задачи проекта

1. Сбор данных о неблагоприятных погодных явлениях, ведущих к экономическим потерям.
2. Предобработка данных: очистка, нормализация и устранение выбросов.
3. Проведение исследовательского анализа данных (EDA) для выявления закономерностей и взаимосвязей.
4. Определение регионов и временных периодов с наибольшими рисками.
5. Подготовка промежуточных и финальных выводов, а также рекомендаций.

2 Выбор задачи и данных

2.1 Выбор задачи

Анализ влияния неблагоприятных погодных явлений на экономические риски является актуальной задачей, поскольку данные явления оказывают значительное влияние на инфраструктуру, сельское хозяйство, транспорт и другие сферы экономики. В рамках проекта мы сосредоточились на оценке:

- Частоты возникновения погодных явлений;
- Интенсивности и продолжительности явлений;
- Географической распределенности явлений.

2.2 Выбор данных

Для анализа использовался массив данных о неблагоприятных погодных явлениях, доступный на сайте ВНИИГМИ-МЦД (Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных). Данные включают сведения о явлениях, наблюдавшихся на территории субъектов Российской Федерации, за период с 1991 по 2023 годы.

Каждая запись в массиве данных содержит следующую информацию:

1. Порядковый номер явления.
2. Дата начала и окончания явления.
3. Количество опасных явлений.
4. Заблаговременность предупреждения.
5. Название явления.
6. Интенсивность явления.
7. Субъект Российской Федерации.
8. Дополнительная информация о территории.

3 Сбор и предобработка данных

3.1 Сбор данных

Данные были загружены с официального сайта ВНИИГМИ-МЦД. Работа с данными велась в Google Colab.

3.2 Обработка данных

1. Загрузка данных: данные загружаются из файла `dataset_input.xlsx`.
2. Обработка данных:
 - Фильтрация и группировка данных по федеральным округам и годам.
 - Вычисление средней интенсивности и продолжительности погодных явлений.
3. Визуализация данных:
 - Построение графиков распределения погодных явлений по федеральным округам.
 - Анализ пар явлений, наблюдаемых одновременно.
4. Экспорт данных: обработанные данные сохраняются в файл `dataset_output.csv`.

4 Исследовательский анализ данных (EDA)

4.1 Знакомство с данными

Набор данных состоит из 3 категориальных, 2 временных и 2 числовых признаков. Это включает в себя такие параметры, как:

- **Event_Name**: название события (например, "Ветер "Метель "Снег"),
- **Region**: регион (например, "Красноярский край "Сахалинская область"),
- **Federal_District**: федеральный округ (например, "Сибирский "Дальневосточный"),
- **Start_Date** и **End_Date**: даты начала и конца события,
- **Event_Intensity** и **Duration**: числовые признаки, отражающие интенсивность и продолжительность события.

4.2 Числовые признаки

4.2.1 Распределение и статистические характеристики

- **Event_Intensity:**

- Среднее значение: 58.39,
- Стандартное отклонение: 118.94,
- Минимальное значение: -65 ,
- Максимальное значение: 1244,
- Квартили: 25% — 25, медиана — 31, 75% — 50.

- **Duration:**

- Среднее значение: 2.16,
- Стандартное отклонение: 8.44,
- Минимальное значение: 0,
- Максимальное значение: 367,
- Квартили: 25% — 0, медиана — 1, 75% — 2.

Event_Intensity имеет большое стандартное отклонение, что указывает на наличие выбросов (максимальное значение 1244). Медиана Event_Intensity (31) значительно ниже среднего значения (58.39), что говорит о возможной асимметрии распределения. Duration имеет более симметричное распределение, так как медиана (1) близка к среднему значению (2.16).

4.2.2 Выбросы

Для анализа выбросов были построены *boxplots*, которые показали наличие выбросов, особенно в столбце Event_Intensity.

4.2.3 Ковариация

Ковариация между числовыми признаками составляет 94.897204. Это указывает на положительную, но относительно слабую связь между Event_Intensity и Duration.

4.3 Временные признаки

4.3.1 Распределение и статистические характеристики

- **Start _Date:**
 - Среднее значение: 2009-04-20 20:45:59,
 - Минимальное значение: 1991-01-01 00:00:00,
 - Квартили: 25% — 2002-06-30 00:00:00, медиана — 2010-03-18 00:00:00, 75% — 2016-12-09 00:00:00,
 - Максимальное значение: 2023-12-31 00:00:00.
- **End _Date:**
 - Среднее значение: 2009-04-23 00:39:43,
 - Минимальное значение: 1991-01-03 00:00:00,
 - Квартили: 25% — 2002-06-30 06:00:00, медиана — 2010-03-19 00:00:00, 75% — 2016-12-10 18:00:00,
 - Максимальное значение: 2024-01-08 00:00:00.

Данные охватывают период с 1991 по 2024 год, что позволяет анализировать долгосрочные тренды. Медиана и quartили указывают на то, что большинство событий происходит в весенне-летний период.

4.4 Категориальные признаки

4.4.1 Распределение и статистические характеристики

- **Event _Name:**
 - Количество уникальных значений: 27,
 - Наиболее частое значение: "Ветер"(3239 раз).
- **Region:**
 - Количество уникальных значений: 82,
 - Наиболее частое значение: "Краснодарский край"(490 раз).
- **Federal _District:**
 - Количество уникальных значений: 8,
 - Наиболее частое значение: "Сибирский"(2533 раза).

Наиболее распространённым событием является "Ветер". Наиболее часто встречающийся регион — Краснодарский край. Среди федеральных округов чаще всего встречается Сибирский округ.

4.5 Корреляция

4.5.1 Корреляция Пирсона

- Сильная положительная корреляция между Start_Date и End_Date (0.999985).
- Слабая положительная корреляция между Event_Intensity и Duration (0.094575).
- Слабая отрицательная корреляция между Event_Name и Event_Intensity (−0.095480).

4.5.2 Корреляция Спирмена

- Сильная положительная корреляция между Event_Name и Event_Intensity (0.347744).
- Слабая отрицательная корреляция между Event_Intensity и Duration (−0.097664).

4.5.3 Phik

- Очень сильная положительная корреляция между Start_Date и End_Date (1.000000).
- Сильная положительная корреляция между Event_Name и Event_Intensity (0.785173).
- Сильная положительная корреляция между Region и Federal_District (1.000000).

4.6 Анализ средней интенсивности и продолжительности событий по федеральным округам

Средняя продолжительность и интенсивность погодных явлений варьируются между федеральными округами. Например:

- Центральный округ: аномально жаркая погода имеет среднюю продолжительность 26.65 дней.

- **Дальневосточный округ:** аналогичное явление длится 6.75 дней.
- **Метель:** продолжительность в Сибирском округе составляет 1.33 дня, а в Северо-Западном — 0.60 дня.

4.7 Анализ экономических рисков от погодных явлений

Регионы с повышенным риском:

- **Дальневосточный округ:** высокие риски в 2005, 2009, 2018 и 2023 годах.
- **Приволжский округ:** высокие риски в 1998, 2007, 2010 и 2015 годах.
- **Сибирский округ:** высокие риски в 1993, 1998, 2005 и 2015 годах.

4.8 Определение пар явлений

Наиболее частые сочетания:

- Ветер и Дождь: 264 раза.
- Ветер и Метель: 252 раза.
- Ветер и Град: 230 раз.

5 Промежуточные и финальные выводы

5.1 Промежуточные выводы

На основе проведенного анализа можно выделить следующие наблюдения:

- **Региональные риски:** регионы, наиболее подверженные экстремальным погодным явлениям, включают Сибирь, Урал и Северный Кавказ. Эти регионы демонстрируют высокую частоту явлений, таких как сильные ветры, изменения температуры и осадки.
- Средняя продолжительность явлений может значительно различаться между федеральными округами, что важно для оценки потенциальных экономических потерь.

- **Временные риски:** погодные явления чаще всего происходят в зимние месяцы, но лето также связано с определенными рисками из-за высокой интенсивности явлений. Построенные распределения по годам показывают, что наиболее рискованные периоды связаны с резкими изменениями температуры и осадками в зимнее время.

5.2 Финальные выводы и рекомендации

Рекомендации:

- Для снижения экономических потерь рекомендуется усилить системы предупреждения о экстремальных погодных явлениях, особенно в регионах с высоким риском, таких как Сибирь и Северный Кавказ.
- Анализ сочетаний явлений, которые часто наблюдаются одновременно, позволит улучшить целенаправленные меры для уменьшения ущерба, особенно в уязвимых регионах.