**ВТУ Св. Св. Кирил и Методий**

**По дисциплината:**

## ,,Извличане на информация”

**проф. дн Цветанка Любомирова Георгиева-Трифонова**

**Изготвил:Синтия Емилова Каратотева**

**Спец.,,Медии и пиар комуникация”**

**задочна форма**

**Фак.номер:2402013921**

**Тема на курсова работа:**

**Обработка и анализ на бизнес информация.**

**1. Набор от данни**

• **Избиране и изтегляне на данни:**

Изберете подходящ набор от данни от онлайн платформи с отворен достъп, например:

• [Kaggle](https://www.kaggle.com/datasets)

• [Open Data Portal EU](https://data.europa.eu/)

• [data.gov](https://www.data.gov/)

• Примерен избор: **“Global COVID-19 Dataset”** от Kaggle.

• **Описание на набора от данни:**

• **Предназначение:** Анализ на разпространението на COVID-19 по държави и региони, за да се визуализират тенденциите в новите случаи и да се разберат ключовите моменти в пандемията.

• **Структура:**

Наборът от данни съдържа следните колони:

• Date: Дата на отчет (в ISO формат).

• Country: Държава, за която се отнасят данните.

• New Cases: Брой нови регистрирани случаи за деня.

• Total Cases: Общо случаи до този ден.

• Population: Население на държавата.

• **Формат:** CSV (Comma-Separated Values).

• **Адрес за достъп:**

Например: [COVID-19 Dataset на Kaggle](https://www.kaggle.com/datasets).

**2. Пречистване на данни**

• **Инструменти:** Google Таблици и OpenRefine.

• **Етапи на пречистване:**

1. **Импортиране:**

• Зареждане на CSV файла в Google Таблици.

• Проверка за правилно разпознаване на колоните.

2. **Премахване на липсващи стойности:**

• Откриване на редове с празни клетки (напр. липсващи данни за нови случаи).

• Изтриване или заместване с подходяща стойност (например “0” за липсващи нови случаи).

3. **Стандартизиране на данни:**

• Корекция на имената на държавите (например “USA” → “United States”).

• Уеднаквяване на формата на датите (например “01/01/2020” → “2020-01-01”).

4. **Форматиране:**

• Преобразуване на числови стойности, за да са в правилния формат (например стойности в текстови формат → числа).

• **Примерен резултат от пречистването:**

• **Преди:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Country | New cases | Total cases | Population |
| 1/1/2020 | USA | 100 | 100 | N/A |
| 2/1/2020 | Germany | N/A | 200 | 83000000 |

* **След:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Country | New cases | Total cases | Population |
| 1/1/2020 | United States | 100 | 100 | 331000000 |
| 2/1/2020 | Germany | 0 | 200 | 83000000 |

**3. Филтриране, обобщаване, групиране**

• **Инструмент:** Google Таблици (Pivot Tables).

• **Процес:**

1. **Създаване на осева таблица:**

• Вмъкване на осева таблица от менюто в Google Таблици.

• **Редове:** Добавяне на колоната Country (по държави).

• **Колони:** Добавяне на колоната Date (групирана по месеци или години).

• **Стойности:**

• New Cases: Сумиране на новите случаи за всеки период.

• Total Cases: Изчисляване на максимална стойност (общо случаи до момента).

2. **Филтриране:**

• Показване на данни само за определени държави или региони.

• Пример: Филтриране на Европа.

3. **Обобщаване:**

• Изчисляване на средни стойности за новите случаи по държави.

4. **Групиране:**

• Събиране на данни по континенти (групиране на държавите).

• **Примерен резултат:**

Pivot таблица с новите случаи за Германия през 2021:

|  |  |  |
| --- | --- | --- |
| Месец | Нови случаи | Общо случаи |
| Месец Януари | 50 000 | 2 000 000 |
| Месец Февруари | 40 000 | 2 040 000 |

**4. Визуализиране на данни**

• **Инструмент:** Google Data Studio.

• **Процедура:**

1. **Свързване на данните:**

• Импортирайте Google Таблици в Google Data Studio.

• Проверете съвместимостта на колоните (напр. дати, числови стойности).

2. **Създаване на визуализации:**

• **Линейна графика:** Показва растежа на общите случаи по държави.

• **Карта:** Географска визуализация, показваща броя случаи по държави.

• **Стълбовидна диаграма:** Сравнява средните нови случаи по континенти.

• **Примерен резултат:**

• **Линейна графика:**

Растеж на COVID-19 случаите в Германия (месечно).

• **Карта:**

Световна карта, където държавите са оцветени според броя на общите случаи (по-наситени цветове за повече случаи).

**5. Споделяне на данни**

• **Инструменти:**

• **Google Таблици:** Създаване на публична връзка за достъп до таблицата.

• **Google Data Studio:** Споделяне на интерактивен отчет чрез публичен линк или вградено представяне в уебсайт.

• **GitHub:** Качване на CSV файловете и документацията за проекта в публично хранилище.

• **Описание:**

• **Предназначение:** Улесняване на достъпа до обработените и визуализираните данни за екипи или широката публика.

• **Обхват:** Всички данни са публично достъпни, с възможност за преглед и изтегляне.

Проектът демонстрира как данните могат да се трансформират от суров вид в структуриран и анализируем формат. От събирането и пречистването до визуализацията и споделянето, всеки етап играе ключова роля за извличане на ценни инсайти.