МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

імені ІГОРЯ СІКОРСЬКОГО"

Факультет прикладної математики

Кафедра програмного забезпечення комп'ютерних систем

Лабораторна робота № 1

з дисципліни "Технології оброблення великих даних

<table>
<tr><td>Виконав</td><td>Зарахована</td></tr>
<tr><td>студент III курсу</td><td>"_____" "_____" 20____ р.</td></tr>
<tr><td>групи КП-82</td><td>викладачем</td></tr>
<tr><td>Мельничук Олексій Геннадійович</td><td>Олещенко Любов'ю Михайлівною</td></tr>
<tr><td>(прізвище, ім'я, по батькові)</td><td>(прізвище, ім'я, по батькові)</td></tr>
</table>

варіант № 12

Київ 2021

Part 1: Open Government Partnership
- Brazil, Ukraine, Australia, Albania etc.
- The government transparency increases overall trust from people and allows for better economic relations.

Part 2: Open Data Websites
- 192,352 datasets
- CSV, JSON, RDF, XML

Part 3: Data Visualization
Knoema:
- Viz of the day: Share of Population Vaccinated against COVID-19 by Country
- Sources: https://www.ourworldindata.org/coronavirus-source-data

Gapminder:
- 'TED Talk: How not to be ignorant about the world' is a great speech that taught me about news bias and how public opinion can influence scientific advancements
- 'Population growth explained with IKEA boxes' is a peculiar speech that explains just how exactly can you stop population growth by ending poverty

Gapminder World:
- Life expectancy began to steadily increase after 1900's and it grows at what appears to be exponential rate
- In 1921 the US was the first country to exceed life expectancy of 60. Germany started to catch up in 1925, and Japan's rate started to skyrocket from 1945 and began to even out from 1973.

Reflection:
- I used https://www.awwwards.com as my website for data visualization. I found data about COVID spread rates and performed some visualizations to visually evaluate data. What I've learned is that basically charts are quite handy for quick data estimation and for making approximate predictions based on current information.

Part 4: Explore Terms of Service Agreements
A. I have accounts at LinkedIn, Spotify and YouTube and I almost never read ToS. However there is a website that shows shortened and concise versions.
B. The Data Use Policy is a compulsory legal disclosure of how a website operator collects, retains and shares personally identifiable information.
C. Privacy settings are "the part of a social networking website, internet browser, piece of software, etc. that allows you to control who sees information about you.
D. Security policy is a definition of what it means to be secure for a system, organization or other entity.
E. In Ukraine, there is no specific legislation regulating online privacy in Ukraine. However, the Data Protection Law applies to the extent online activities involve the processing of personal data.
F. The provider has complete control over the data you upload and download.
G. Some services may delete it, but majority of them store it indefinitely.

Part 5: Do You Know What You Signed Up For?

A. Use that service that gives you concise TOS. If it doesn't work, search for some keywords.
B. Use different passwords.

Part 6: Explore and Download the Data

- 3,452,081 rows and 6 columns
- StationName, StationLocation, DateTime, RecordId, RoadSurfaceTemperature, AirTemperature
- 296 Mb

Part 7: Explore the Limitations of a Spreadsheet

- Warning message: Possible data loss.
- How will the message impede my analysis: I have incomplete data on the subject, therefore some predictions and measurements may prove to be incorrect and very approximated.
- Maximum amount of rows: 1,048,576
- Maximum amount of rows in spreadsheet of my choice: Depends on amount of data
- The Road Weather Information Stations dataset has exceeded the maximum amount of rows possible in Excel. This would explain why there was a data loss error while importing the file.

Part 8: Дослідження порталу відкритих даних України

- Будівництво, Держава, Екологія, Економіка та бізнес, Земля, Молодь і спорт, Освіта і культура, Охорона здоров'я, Податки, Сільське господарство, Соціальний захист, Стандарти, Транспорт, Фінанси, Юстиція
- XLSX, ZIP, CSV, JSON
- Розположення деяких посилань