CMPS 360 L51

Data Science Fundamentals

Spring 2025

Dr. Tamer Elsayed

Project Part B: Data Science in Action

Genocide in Gaza: Telling the Story through Data Science

Ganna - 202205763

Atrab - 000000000

Non - 000000000

Fatma – 000000000

Contents

[Introduction 3](#_Toc197205918)

[Preprocessing 3](#_Toc197205919)

[Datasets 3](#_Toc197205920)

[Daily Casualties – Gaza 4](#_Toc197205921)

[Daily Casualties – West Bank 4](#_Toc197205922)

[Killed in Gaza 5](#_Toc197205923)

[Infrastructure Damaged 5](#_Toc197205924)

# Introduction

In this part of the project, we are going to explore and analyze the provided data about the ongoing genocide in Gaza. We will try to find interesting observations and explore potential associations. We will also perform hypothesis testing.

# Preprocessing

* Converted the PDF list in the "killed-in-gaza\_moh\_2024-09-21.pdf" file to CSV.
  + Used an online tool that converts pdf to xlsx then used that to create csv (“killed-in-gaza.csv”).
  + <https://www.ilovepdf.com/pdf_to_excel>
* Converted the JSON list in the "infrastructure-damaged.json" file to CSV.
  + Used an online tool that converts json to xlsx then used that to create csv (“infrastructure-damaged.csv”).
  + <https://data.page/json/csv>

# Datasets

We have 4 different data sets, providing different data related to the genocide:

* Daily Casualties - Gaza
* Daily Casualties – West Bank
* Killed in Gaza
* Infrastructure damaged

All are initially csv files that we convert to tables as a first step.

## Daily Casualties – Gaza

This dataset provides daily values for those killed and injured in the Gaza Strip since October 7th, 2023. There are 551 days of reports from 2023-10-07 to 2025-04-09.

Table name: **gaza\_daily\_casualties**

|  |  |
| --- | --- |
| **field name** | **value** |
|  |  |
| report\_date | date in YYYY-MM-DD |
| report\_source | one of mohtel, gmotel, unocha or missing |
| report\_period | one of 24, 48, or 0, representing the hours length of the reporting period |
| ext\_massacres\_cum | extrapolated cumulative number of multiple fatality events |
| killed | killed persons on the given report date |
| killed\_cum | cumulative number of killed persons to the report date |
| ext\_killed | same as killed but extrapolated |
| ext\_killed\_cum | same as killed\_cum but extrapolated |
| ext\_killed\_children\_cum | extrapolated cumulative number of children killed to the report date |
| ext\_killed\_women\_cum | extrapolated cumulative number of women killed to the report date |
| injured\_cum | cumulative number of injured persons to the report date |
| ext\_injured | injured persons on the given report date |
| ext\_injured\_cum | cumulative number of injured persons to the report date |
| ext\_civdef\_killed\_cum | extrapolated cumulative number of emergency services killed to date |
| med\_killed\_cum | cumulative number of medical personnel killed to the report date |
| ext\_med\_killed\_cum | extrapolated cumulative number of medical personnel killed to the report date |
| press\_killed\_cum | cumulative number of journalists killed to the report date |
| ext\_press\_killed\_cum | extrapolated cumulative number of journalists killed to the report date |

## Daily Casualties – West Bank

This dataset provides daily values for those killed and injured in the West Bank since October 7th, 2023. There are currently 551 days of reports from 2023-10-07 to 2025-04-09.

Table name: **west\_bank\_daiy\_casualties**

|  |  |
| --- | --- |
| **field name** | **value** |
| report\_date | date in YYYY-MM-DD |
| verified.killed | killed persons on the given report date |
| verified.killed\_cum | cumulative number of confirmed killed persons to the report date |
| verified.injured | injured persons on the given report date |
| verified.injured\_cum | cumulative number of injured persons to the report date |
| verified.killed\_children | number of children killed on the given report date |
| verified.killed\_children\_cum | cumulative number of children killed to the report date |
| verified.injured\_children | number of children killed on the given report date |
| verified.injured\_children\_cum | cumulative number of children injured to the report date |
| killed\_cum | same as verified.killed\_cum but yet to be independently verified |
| killed\_children\_cum | same as verified.killed\_children\_cum but yet to be independently verified |
| injured\_cum | same as verified.injured\_cum but yet to be independently verified (see below) |
| injured\_children\_cum | same as verified.injured\_children\_cum but yet to be independently verified |
| settler\_attacks\_cum | cumulative number of attacks by settlers on civilians to the report date |
| flash\_source | either un or fill, see [March 25 update](https://data.techforpalestine.org/updates/west-bank-frequency/) for more detail |

## Killed in Gaza

## Infrastructure Damaged