

CRIME ANALYSIS IN CHICAGO IN THE YEARS 2018-2022

TAKING INTO ACCOUNT ECONOMIC AND DEMOGRAPHIC FACTORS

Data Warehouses 2025

Dominic Czech (272725)

Contents

Stage 1 – Project Development Scope	3
1. Project Title	3
2. Characteristics of the problem domain.....	3
3. The purpose of the project and the scope of the analysis.....	3
Project objective	3
Scope of analysis	4
4. Data Sources	5
5. Scope of Data	6
Crimes 2001 to Present	6
ACS_5_Year_Data_By_Ward.....	7
IUCR Codes.....	9
6. Entity/Class Type Definition.....	10
7. Examples of multivariate analyses	10
8. Creation of the Database	12
Stage 2 – ETL Process	14
1. Preparation of Source Data	14
2. Data Warehouse Architecture.....	14
3. ETL Process Structure (SSIS)	15
1. Delete Tables	15
2. Create Tables.....	16
3. Load Dim_TimeOfDay.....	18
4. Load Dim_Date	18
5. Load Dim_WardDemographics.....	19
6. Load Dim_IUCR.....	20
7. Load Dim_Location	21
8. Load Fact_Crime.....	22
9. Constraints (References)	24
10. Success Message	26

11.	Error Handling	26
Step 3 – OLAP Cube in SSAS.....		27
1. Stage Goal Description		27
2. Cube Structure		27
3. Dimensions		28
Dim_Date:		28
Dim_TimeOfDay.....		29
Dim_Location		29
Dim_WardDemographics.....		29
Dim_IUCR.....		30
Dim_CrimeFlags (degenerate dimension):		30
4. Analyses – Stage 3 (implementation of point 7 of Stage 1)		31
Analysis 1: Distribution of crime categories by dominant ethnic group.....		31
Analysis 2: Types of crimes vs. ward economic status		36
Analysis 3: Crime in Wards with Young Populations		41
Analysis 4: Arrest Effectiveness by Economic Status.....		43
Analysis 5: Domestic crime rates and age and ethnic profiles		48
Analysis 6: Crime Trend by Ward Economic Status		49
Analysis 7: Crimes by day of the week and location.....		51
Analysis 8: Number of crimes by type and outcome (arrest)		53
Analysis 9: Arrest effectiveness over the years.....		58
Analysis 10: Crime Category by Time of Day		63
5. KPI + Reports.....		66
6. Conclusions from Data Analysis:		67

Stage 1 – Project Development Scope

1. Project Title

Analysis of crime in Chicago with regard to economic and demographic factors and UCR classification

2. Characteristics of the problem domain

The problem area of the project is **the analysis of urban crime** in the context of **socioeconomic conditions** in the city of **Chicago** in the years **2018–2022**.

Modern cities, especially metropolises such as Chicago, face a high level of variability in crime depending on location, time, and the social and economic conditions of the population living there.

The City of Chicago publishes a wealth of public data on **crime** (e.g., robberies, thefts, assaults, murders) that includes information **on the type of crime, location, time, and arrest of the perpetrator**. The mere compilation of this information can provide valuable statistics, but combining it with additional dimensions—e.g., **income, demographics, and ethnicity**—

allows for a more complete picture of the conditions that contribute to crime.

The goal is to examine how factors such as age, gender, income, and ethnicity of the population in different areas of Chicago may correlate with crime rates and the effectiveness of law enforcement. Such analysis can help identify areas requiring social intervention, changes in security policies, or better allocation of police and city resources.

3. The purpose of the project and the scope of the analysis

The aim of the project

The main objective of the project is **to develop a data warehouse** that will enable analysis of crime in Chicago from 2018 to 2022, taking into account socioeconomic and demographic data from the American Community Survey (ACS) and other public sources. The project involves creating **an OLAP cube** that will enable advanced multidimensional analyses in order to:

- identification of **crime trends** depending on location (district/ward), time (year, month) and type of crime,
- finding the relationship between crime and the demographic structure, income and ethnic origin of residents,

- support **the decision-making process** for city services, social policy planners and city authorities,
- creating a basis for **building predictive models** in the future that could predict an increase in the crime threat in specific areas.

With data warehousing and OLAP analytics, end users will be able to quickly get answers to questions like:

- Do higher incomes in a given area correlate with lower crime?
- What types of crimes predominate in neighborhoods with a predominantly young population?
- Are certain ethnic groups more likely to be victims or perpetrators of certain crimes?
- How has the number of arrests changed over the years and is this related to demographic data?

Scope of analysis

The scope of the analyses will include data from **2018–2022** to ensure consistency with the annual scope of the ACS 2023 data (which is a 5-year aggregate). The project assumes the analysis of the following aspects:

1. Time:

- Year, month, day of the week, time of day – allows you to analyze seasonality crimes.

2. Location:

- Ward (political district),
- Community Area (geographic area),
- District, Police District (District, Beat),
- Geographic coordinates (for possible visualization on maps).

3. Types of crimes:

- Classification by type of crime (e.g. robbery, assault, burglary),
- Detailed description of the crime (Description, IUCR),
- Information as to whether an arrest has occurred.

4. Socio-economic factors:

- Income ranges of residents in a given Ward,
- Population distribution by age and gender,
- Racial and ethnic structure of the population,
- Size of the general population.

5. Analytical measures (facts):

- Number of crimes,
- Number of crimes ending with arrest,
- The share of a given type of crime in relation to all crimes in the area,
- Number of crimes per 1,000 inhabitants (non-additive measure).

4. Data Sources

No. .	File	Type	Number of records	Size [MB]	Description
1.	Crimes 2001 to Present (https://catalog.data.gov/dataset/crimes-2001-to-present)	csv 83	15520 ~1850		Chicago crime data from 2001- 2025. For the purpose of analyzing socio- economic factors ch in analysis the year range 2018- will be used 2022.
2.	ACS_5_Year_Data_by_Ward.csv (https://catalog.data.gov/dataset/acs-5-year-data-by-ward)	csv	51	0.008	5-Year Survey Results summarized in 2023. Data economic- Demographics for Wards in Chicago (2018-2022 period).
3.	Chicago Police Department - Illinois Uniform Crime Reporting (IUCR) Codes	Csv	410	0.02	Description of Illinois crime codes. Contains

					main and additional description of the crime.
--	--	--	--	--	---

5. Scope of Data

Crimes 2001 to Present

No.	Attribute	Data type	Value range	Comments - quality assessment
1.	ID	INT	634-13837364	Unique data, 0% missing, fact table primary key.
2.	Case Number	STRING 6-9 characters		Investigation ID (useful for multiple crimes in one case). 0% missing.
3.	Date	DATE	Years 2001-2025	Data complete (0% missing). Saved in the format MM/DD/YYYY HH:MM:SS XM.
4.	Block	STRING 4-38 characters		Block address, 0% defects, repeats possible.
5.	IUCR	STRING 4 characters		Uniform Crime Reporting Classification Code, 0% missing.
6.	Primary Type	STRING 5-33 characters		Major crime category, e.g. THEFT, ASSAULT. 0% deficiencies.
7.	Description	STRING		Detailed description of the event. 0% deficiencies.
8.	Location Description	STRING		Location of the incident, 0.17% missing – consider imputation or category “Unknown”.
9.	Arrest	BOOLEAN True/False		Was the offender apprehended (0% deficiencies).
10.	Domestic	BOOLEAN True/False		Was the crime domestic (0% deficiencies).
11.	Beat	INT	111 – 2535	Patrol unit code (0% missing).
12.	District	FLOAT	1.0-31.0	Police district, 0% defects, real number - may require conversion to INT.

13.	Ward	FLOAT	1.0-50.0	Administrative district, 7.39% deficiencies – replenishment should be considered.
14.	Community Area	FLOAT	0.0-77.0	Geographic unit, 7.38% missing – similar to Ward, imputation needed.
15.	FBI CODE	STRING 2-3 characters		FBI Classification - 0% Deficiencies
16.	X Coordinate Float		0.0-125119.00	Coordinates X – 4.06% missing.
17.	Y Coordinate Float		0.0-1951622.00	Y coordinates – 4.06% missing.
18.	Year	INT	2001-2025	Unnecessary, can be extracted from the date, 0% deficiencies.
19.	Updated On Date		Years 2001-2025	Record update date (0% missing).
20.	Latitude	Float	36,619 – 42,022	Latitude 4.06% missing.
21.	Longitude	Float	-91,687 - -87,525	Longitude 4.06% missing.
22.	Location	STRING 23-29 characters		Glued coordinates (lat, long). 4.06% missing

ACS_5_Year_Data_By_Ward

No.	Attribute	Data Type	Value range Notes	
1,	ACS YEAR	INT	2023	Data aggregated in 2023, collected in years 2018-2022. Column unnecessary
2.	WARD INT Under \$25,000		1-50	Ward Number.
3.	INT		0-11000	Number households with a total annual income less than \$25,000.
4.	\$25,000 \$49,999 this	INT	0-11000	Number households with a total

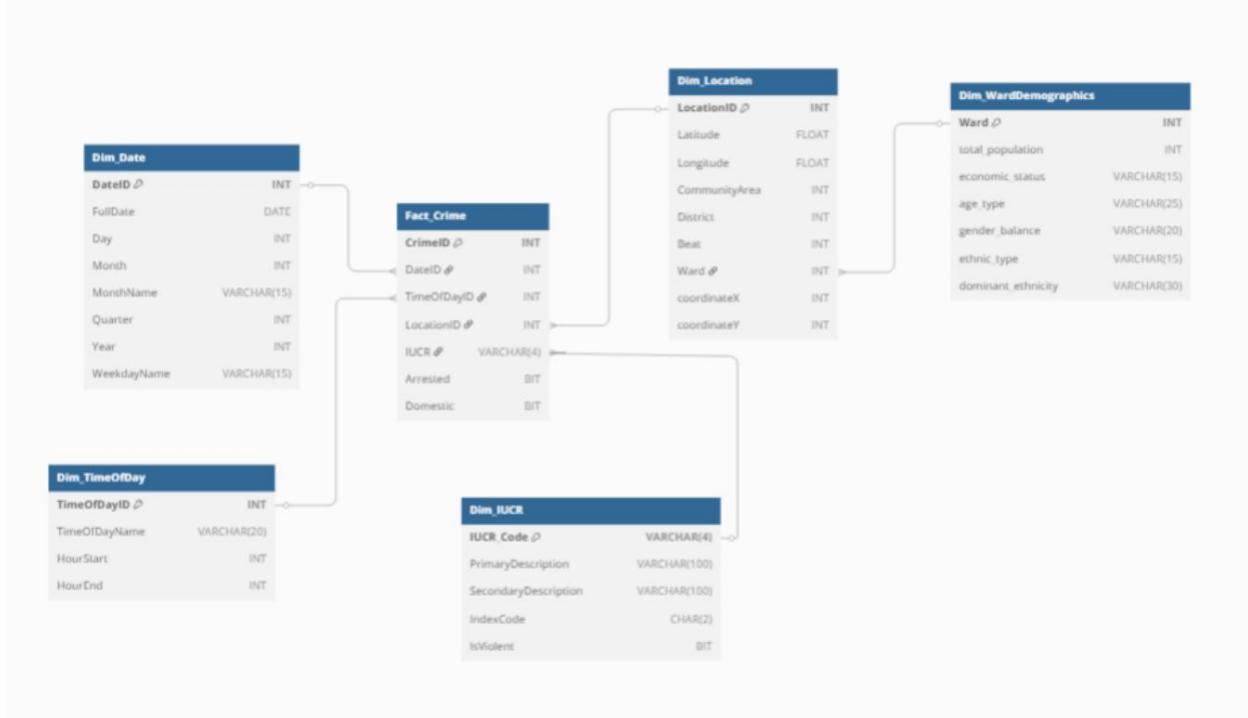
				annual income \$25,000 - \$49,999.
5.	\$50,000 this \$74,999	INT	0 -11000	Number of households with combined annual income of \$50,000 - \$74,999.
6.	\$75,000 this \$125,000	INT	0 -11000	Number households with a total annual income \$75,000 - \$125,000.
7.	\$125,000+	INT	0 -11000	Number of households with combined annual income greater than \$125,000.
8 -13	Male 0 to 17, 18 to 24, 25 to 34, 35 to 49, 50 to 65, 65+	INT	0 -15000	Number of men in a given age group
14 -19	Female 0 to 17, 18 to 24, 25 to 34, 35 to 49, 50 to 65, 65+	INT	0 -15000	Number of women in given age range.
20.	Total Population	INT	0 -75000	Population in Ward.
21.-27.	White, Black or African, American Indian or Alaska Native, Asian, Native Hawaiian or	INT	0 -20000	Population from a given ethnic group in Ward.

	Pacific Islander, Other Flares, Multiracial			
28.	White not Hispanic or Latino	INT	0-50000	Population of white skin and not of Hispanic descent.
29.	Hispanic or Latino INT		0-50000	Population of Hispanic origin.
30.	RecordID	STRING	2023_XX	Combined value of the year with number Ward, column unnecessary.

IUCR Codes

No.	Attribute	Data type	Value range	Comments - quality assessment
1.	IUCR	STRING	4 characters	Crime type code.
2.	Primary Description	STRING	Up to 30 characters	Main type of crime (e.g. HOMICIDE, THEFT, ASSAULT)
3.	Secondary Description	STRING	Up to 100 characters.	A more detailed description of the crime.
4.	Index Code	STRING	I/N	Whether the crime is an "Index Crime" (or prosecuted by the FBI).
5.	Active	BOOLEAN	True/False	Is the crime type still active?

6. Definition of Entity Types/Classes



Additive (calculated) measures:

- ArrestedCount
- DomesticCount
- CrimeCount

Non-additive (calculated) measures:

- CrimesPer1000Inhabitants
- PercentArrested

7. Examples of multivariate analyses

1. Number of crimes by dominant ethnic group

- **Dimensions:** dominant_ethnicity, PrimaryDescription (from Dim_IUCR)
- **Measure:** CrimeCount
- **Objective:** To identify cultural/systemic differences in the distribution of crime depending on the dominant ethnic group in the Ward.

2. Number and type of crimes depending on economic status

- **Dimensions:** economic_status, PrimaryDescription
 - **Measures:** CrimeCount, ArrestCount
 - **Objective:** To examine the relationship between the property status of neighborhoods and the types of crimes and the effectiveness of arrests.
-

3. Predominant types of crime in wards with a young population

- **Dimensions:** age_type, PrimaryDescription
 - **Measure:** CrimeCount
 - **Goal:** Assess risks and adapt preventive programs for youth.
-

4. Effectiveness of arrests by economic status

- **Dimensions:** economic_status, PrimaryDescription
 - **Metric:** ArrestRate (calculated: arrested / all)
 - **Objective:** To analyze the inequality in law enforcement between richer and poorer people Wards.
-

5. Domestic crime rates and age and ethnic profile

- **Dimensions:** age_type, dominant_ethnicity
 - **Measure:** DomesticCrimeRate (share of domestic crimes)
 - **Objective:** To reveal potential correlations between age, ethnicity and violence homely.
-

6. Crime Trends in the Poorest Areas

- **Dimensions:** Year, Ward
- **Filter:** economic_status = 'poor'

- **Measure:** CrimeCount
 - **Goal:** Monitor changes in the most economically challenged neighborhoods.
-

7. Number of crimes by day of the week and location

- **Dimensions:** WeekdayName, Latitude, Longitude (from Dim_Location)
 - **Measure:** CrimeCount
 - **Purpose:** Time-space analysis (e.g. for a heatmap in Power BI).
-

8. Number of crimes by type and outcome (arrest)

- **Dimensions:** PrimaryDescription, Arrested
 - **Measure:** CrimeCount
 - **Objective:** To compare the effectiveness of arrests between crime types.
-

9. Effectiveness of arrests by ethnic group

- **Dimensions:** dominant_ethnicity, Year
 - **Measure:** ArrestRate
 - **Objective:** To explore possible differences in the approach of services to different communities.
-

10. Crimes depending on the time of day and age of residents

- **Dimensions:** age_type, TimeOfDayName
 - **Measure:** CrimeCount
 - **Objective:** To identify when and where young communities are most vulnerable.
-

8. Creation of the Database

```
-- Dim_Date  
CREATE TABLE Projekt.Dim_Date ( DateID  
INT, FullDate  
DATE, Day INT,  
Month INT,
```

```

MonthName VARCHAR(15),
Quarter INT,
Year INT,
WeekdayName VARCHAR(15)
);

-- Dim_TimeOfDay
CREATE TABLE Projekt.Dim_TimeOfDay ( TimeOfDayID
INT, TimeOfDayName
VARCHAR(20), -- Night, Morning, Afternoon, Evening HourStart INT, HourEnd INT

);

-- Dim_WardDemographics CREATE
TABLE Projekt.Dim_WardDemographics ( Ward INT, total_population
INT,
economic_status VARCHAR(15),
age_type VARCHAR(25), gender_balance
VARCHAR(20), ethnic_type
VARCHAR(15), dominant_ethnicity
VARCHAR(30)

);

-- Dim_Location CREATE
TABLE Projekt.Dim_Location ( LocationID INT
IDENTITY(1,1) PRIMARY KEY, Latitude FLOAT, Longitude FLOAT,
CommunityArea INT,
District INT, Beat INT,
Ward INT, coordinateX INT,
coordinateY INT

);

-- Dim_IUCR
CREATE TABLE Projekt.Dim_IUCR ( IUCR_Code
VARCHAR(4), PrimaryDescription
VARCHAR(100), SecondaryDescription
VARCHAR(100), IndexCode CHAR(2), IsViolent BIT
--I, N

);

-- Fact_Crime
CREATE TABLE Projekt.Fact_Crime ( CrimeID INT,
DateID INT,
TimeOfDayID INT, LocationID INT, IUCR
VARCHAR(4), Arrested
BIT, Domestic BIT

);

```

Stage 2 – ETL Process

1. Preparation of Source Data

As the analysis concerns, among others, socioeconomic factors collected in 2018-2022, the first step was to filter the source data.

For this purpose I used a Python script library (Pandas library).

```
import pandas as pd

df = pd.read_csv("Crimes_-_2001_to_Present.csv", parse_dates=['Date'], low_memory=False)

df_filtered = df[df['Date'].dt.year.between(2018, 2022)]

df_filtered.to_csv("filtered_crimes.csv", index=False)
```

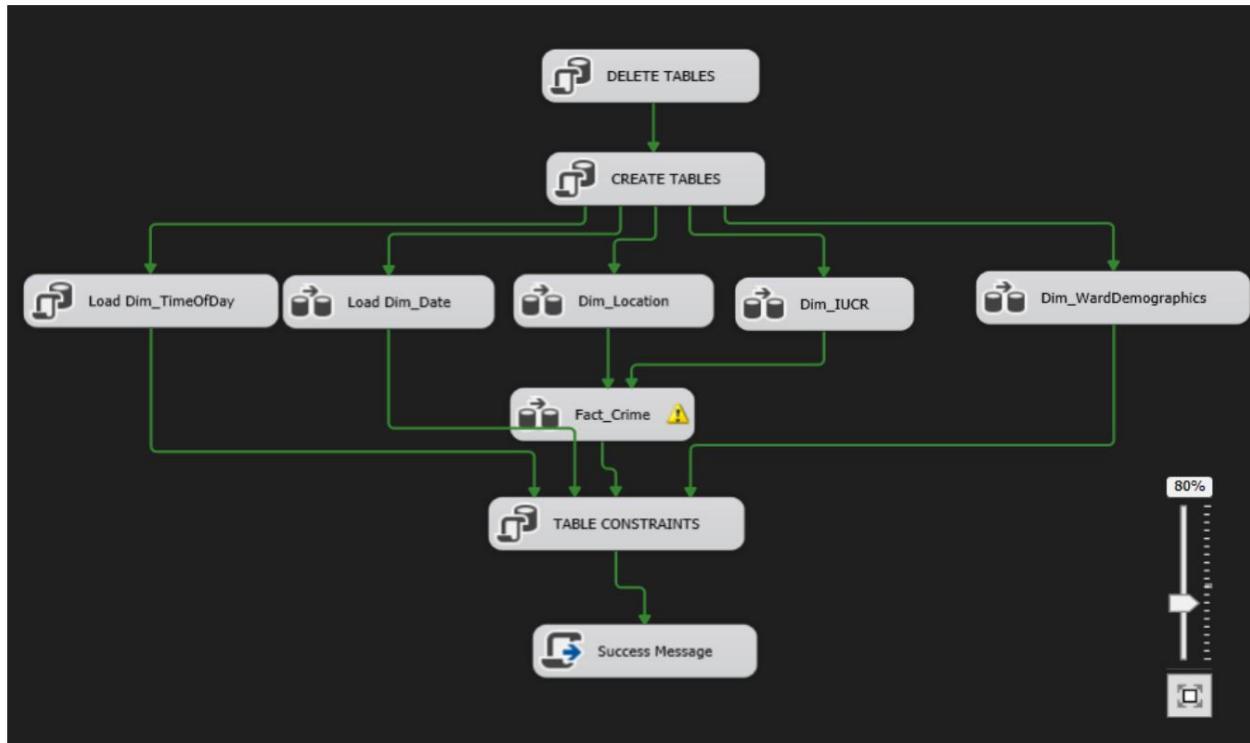
2. Data Warehouse Architecture

The following tables were implemented in the data warehouse:

- **Fact:** Fact_Crime
- **Dimensions:** Dim_Date, Dim_TimeOfDay, Dim_Location, Dim_IUCR,
Dim_WardDemographics

3. ETL Process Structure (SSIS)

The ETL process is divided into **modular SSIS packages**. Each package performs a specific function. All packages are run sequentially and are saved in the SSIS solution "**Projekt_ETL**".



1. Delete Tables

Description:

This task is responsible for cleaning data from target tables in the data warehouse so that the ETL process can be performed from scratch (full data load).

Task type: Execute SQL Task

SQL code:

```

IF EXISTS (
    SELECT *
    FROM INFORMATION_SCHEMA.TABLES
    
```

```

        WHERE TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Fact_Crime'
    )
DROP TABLE Projekt.Fact_Crime; HIM

IF EXISTS
( SELECT *
  FROM INFORMATION_SCHEMA.TABLES WHERE
  TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Dim_Date'
)
DROP TABLE Project.Dim_Date; HIM

IF EXISTS
( SELECT *
  FROM INFORMATION_SCHEMA.TABLES WHERE
  TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Dim_TimeOfDay'
)
DROP TABLE Project.Dim_TimeOfDay; HIM

IF EXISTS
( SELECT *
  FROM INFORMATION_SCHEMA.TABLES WHERE
  TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Dim_Location'
)
DROP TABLE Project.Dim_Location; HIM

IF EXISTS
( SELECT *
  FROM INFORMATION_SCHEMA.TABLES WHERE
  TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Dim_WardDemographics'
)
DROP TABLE Project.Dim_WardDemographics; HIM

IF EXISTS
( SELECT *
  FROM INFORMATION_SCHEMA.TABLES WHERE
  TABLE_SCHEMA = 'Project' AND TABLE_NAME = 'Dim_IUCR'
)
DROP TABLE Projekt.Dim_IUCR; HIM

```

2. Create Tables

Description: The task creates the structure of fact and dimension tables in the Project database, if it has not been created yet. In practice, it can be performed only once - during project initialization.

Task type: Execute SQL Task

SQL – table structures:

```
-- Dim_Date
CREATE TABLE Projekt.Dim_Date ( DateID
    INT, FullDate
    DATE, Day INT,
    Month INT,
    MonthName
    VARCHAR(15), Quarter INT,
    Year INT,
    WeekdayName VARCHAR(15)
);

-- Dim_TimeOfDay
CREATE TABLE Projekt.Dim_TimeOfDay ( TimeOfDayID
    INT, TimeOfDayName
    VARCHAR(20), -- Night, Morning, Afternoon, Evening HourStart INT, HourEnd INT
);

-- Dim_WardDemographics
CREATE TABLE Projekt.Dim_WardDemographics ( Ward
    INT,
    total_population INT,
    economic_status VARCHAR(15),
    age_type VARCHAR(25),
    gender_balance VARCHAR(20),
    ethnic_type VARCHAR(15),
    dominant_ethnicity VARCHAR(30)
);

-- Dim_Location CREATE
TABLE Projekt.Dim_Location ( LocationID INT
    IDENTITY(1,1) PRIMARY KEY, Latitude FLOAT, Longitude FLOAT,
    CommunityArea INT,
    District INT, Beat INT,
    Ward INT, coordinateX INT,
    coordinateY INT
);

-- Dim_IUCR
CREATE TABLE Projekt.Dim_IUCR ( IUCR_Code
    VARCHAR(4), PrimaryDescription
    VARCHAR(100), SecondaryDescription
    VARCHAR(100), IndexCode CHAR(2), IsViolent BIT
    -- e.g. I, II, N
);

-- Fact_Crime
CREATE TABLE Projekt.Fact_Crime ( CrimeID INT,
```

```

    DateID INT,
    TimeOfDayID INT,
    LocationID INT,
    IUCR VARCHAR(4),
    Arrested BIT,
    Domestic BIT
);

```

3. LoadDim_TimeOfDay

Purpose: Manually populate the Dim_TimeOfDay table.

Description: Static data – 4 times of day: Night (0–6), Morning (6–12), Afternoon (12–18), Evening (18–24). Entered as T-SQL Insert.

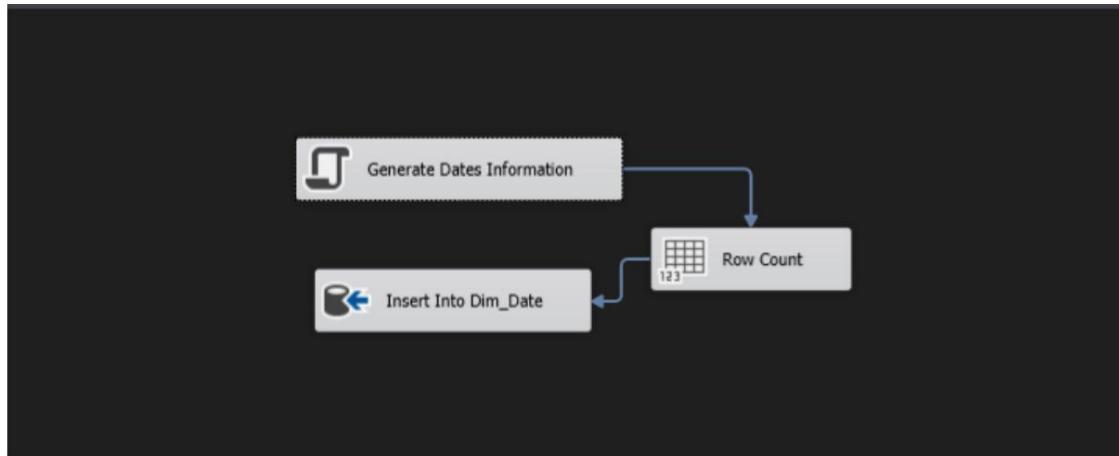
SQL – data population:

```

INSERT INTO Projekt.Dim_TimeOfDay (TimeOfDayID, TimeOfDayName, HourStart, HourEnd)
VALUES
(1, 'Morning', 6, 11),
(2, 'Afternoon', 12, 17),
(3, 'Evening', 18, 23),
(4, 'Night', 0, 5);

```

4. LoadDim_Date



Purpose: Populate the Dim_Date table with the range of dates present in the set.

Description: A C# script was used to generate dates in the range 2018–2022 along with date attributes (day, month, day name, quarter, etc.).

C# script:

```

DateTime startDate = new DateTime(2018, 1, 1);
DateTime endDate = new DateTime(2022, 12, 31);

```

```

for (DateTime date = startDate; date <= endDate; date = date.AddDays(1))
{
    Output0Buffer.AddRow();

    Output0Buffer.FullDate = date;
    Output0Buffer.Day = date.Day;
    Output0Buffer.Month = date.Month;
    Output0Buffer.MonthName = date.ToString("MMMM",
CultureInfo.InvariantCulture);
    Output0Buffer.Quarter = (date.Month - 1) / 3 + 1;
    Output0Buffer.Year = date.Year;
    Output0Buffer.WeekdayName = date.ToString("dddd",
CultureInfo.InvariantCulture);

    // DateID in YYYYMMDD format
    Output0Buffer.DateID = int.Parse(date.ToString("yyyyMMdd"));
}
}
}

```

5. Load Dim_WardDemographics



Source: ACS 5 Year Data by Ward

Goal: Load and transform sociodemographic data into Dim_WardDemographics

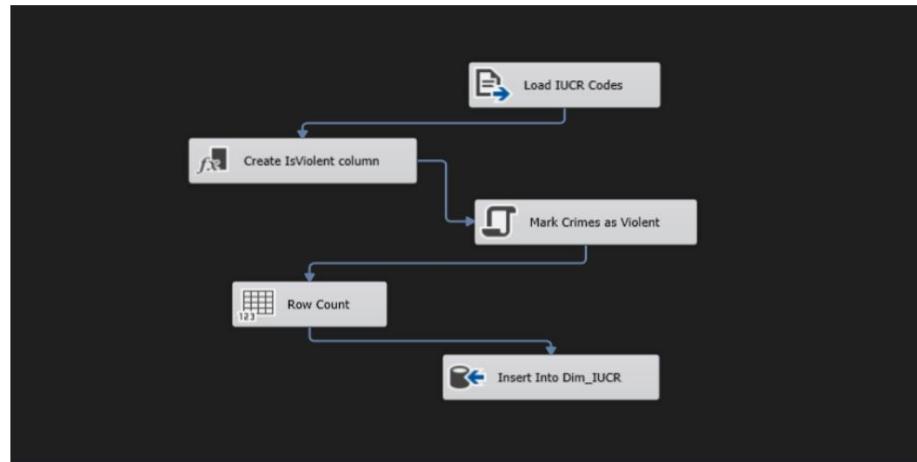
Description: Data classified manually for the project, includes classifications:

- economic_status:
 - o "extreme" – if the income of more than 40% of households is below \$50,000 and the income of more than 40% of households is above \$125,000
 - o "poor" – if the median household income is below \$50,000
 - o "rich" - if the median household income is above \$125,000

about "average" - in the rest of the cases

- age_type
 - young-dominant – if more than 55% of the ward's residents are under 35 years old life
 - elderly-dominant – if more than 35% of the ward's residents are over 50 years old life
 - about balanced – in the remaining cases
- gender_balance
 - male-dominant – if more than 55% of the ward's inhabitants are male
 - female-dominant – if more than 55% of the ward's inhabitants are women
 - mixed – in the rest of the cases
- ethnic_type
 - dominant – if one ethnic group constitutes more than 50% of the population ward
 - mixed - otherwise
- dominant_ethnicity – the largest ethnic group in the ward (e.g. White, Hispanic, Black)

6.LoadDim_IUCR



Source: Chicago Police Department IUCR Codes

Purpose: To retrieve IUCR police codes and classify crimes as violent crimes.

Description: Violent crimes are classified as:

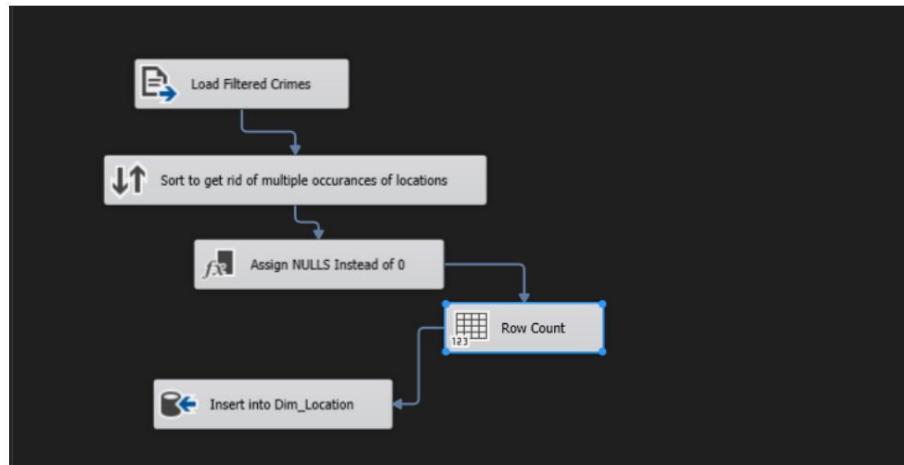
- Murder

- Criminal sexual assault
- Robbery
- Beating
- Ritual crimes
- Assault
- Human Trafficking
- Sexual offence (sexual exploitation of a child/criminal sexual abuse/criminal transmission of HIV/attempted aggravated criminal sexual abuse)
- Crime involving children (child pornography/criminal sexual abuse by a family member/sexual assault by a family member)

7.LoadDim_Location

Source: FilteredCrimes

Purpose: To load unique combinations of geographic coordinates and territorial data (e.g. Community Area, Beat, Ward) into the Dim_Location dimension table, allowing for the analysis of crime from a spatial perspective.



Description:

Due to the lack of Lookup capabilities for float values (Longitude and Latitude), we retain the following location data that allows us to uniquely identify a location at the time of Lookup during the Fact_Crimes data loading process:

- Community Area
- District
- Beat
- Ward

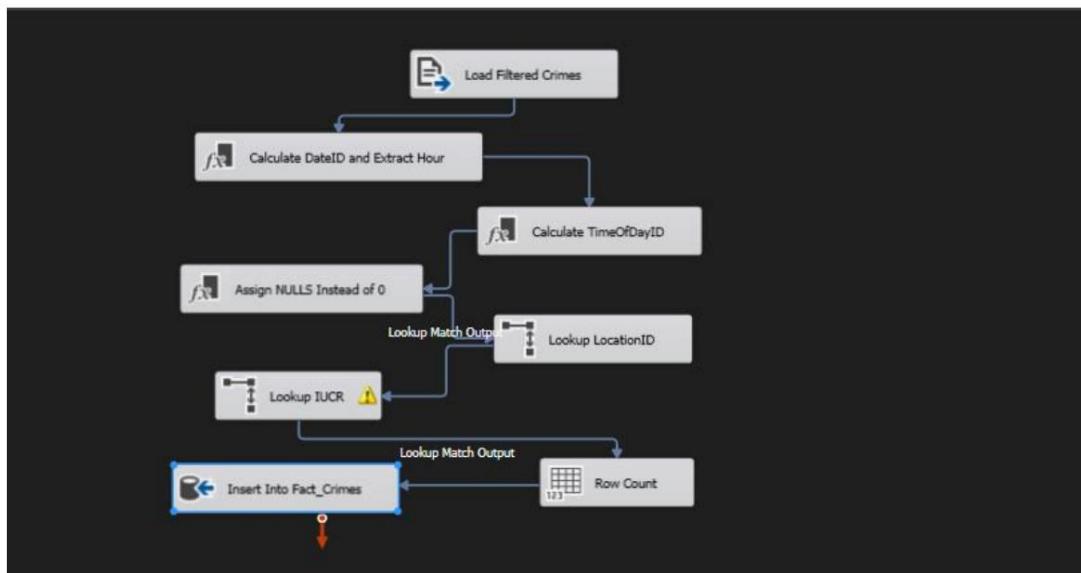
- X Coordinate
- Y Coordinate

And data that will allow us to analyze the spatial nature of crimes:

- Longitude
- Latitude

If there is a 0 in any of the columns (impossible in the Chicago data) – we swap these value to NULL.

8. Load Fact_Crime



Purpose: Loading the Fact_Crime fact table based on data from the source file, joined with the Dim_Date, Dim_TimeOfDay, Dim_Location, Dim_IUCR dimensions.

Processing steps (in SSIS - Data Flow):

1. Loading data from a CSV file

- **Source:** Crimes CSV file (e.g. Crimes_2018_2022.csv)
- **Task:** Flat File Source
- **Description:** Loads all records from a raw dataset, including date, time, location, arrest, etc.

2. Derived Column – extract DateID and Hour

- **Quest:** Derived Column

- **Description:** Extracts the date and time from a Date field in text format and creates:

- o DateID in the format yyyyymmdd

- o Hour as a number from 0 to 23

3. Derived Column – determining TimeOfDayID based on Hour

- **Quest:** Derived Column

- **Description:** For each hour, an appropriate TimeOfDayID is assigned:

- at 1: Morning (6–11)

- at 2: Afternoon (12–17)

- at 3: Evening (18–23)

- at 4: Night (0–5)

4. Derived Column – cleaning up invalid locations (0 -> NULL)

- **Task:** Another **Derived Column** component

- **Description:** Replaces 0 values in localization fields with NULL values that have no analytical significance.

5. Lookup - getting LocationID from Dim_Location

- **Task:** Lookup

- **Description:** Matching the location to the Dim_Location dimension based on:

- o Latitude, Longitude, CommunityArea, District, Ward, Beat, coordinateX, coordinateY

6. Lookup – checking if the code is in IUCR

- **Task:** Lookup •

Description: Matching the IUCR code to the existing ones in the Dim_IUCR table. If match does not exist, row is rejected.

7. Selecting and mapping data to Fact_Crime

1. Output Columns:

- about CrimeID (from file)

- about DateID (from Derived Column)

- o TimeOfDayID (from Derived Column)

- about LocationID (from Lookup)

- about IUCR (from file)

- about Arrested (from file)

- about Domestic (from file)

9. Constraints (References)

Goal: To assign correct constraints (PK, FK) to tables filled with data.

Task: Execute SQL Task

SQL:

```
-- PRIMARY KEYS
```

```
-- Dim_Date IF
```

```
NOT EXISTS
```

```
( SELECT * FROM sys.key_constraints
  WHERE name = 'PK_Dim_Date' AND parent_object_id = OBJECT_ID('Projekt.Dim_Date')
)
```

```
ALTER TABLE Projekt.Dim_Date ADD
```

```
CONSTRAINT PK_Dim_Date PRIMARY KEY (DateID);
```

```
-- Dim_TimeOfDay IF
```

```
NOT EXISTS
```

```
( SELECT * FROM sys.key_constraints
  WHERE name = 'PK_Dim_TimeOfDay' AND parent_object_id =
OBJECT_ID('Projekt.Dim_TimeOfDay') )
```

```
ALTER TABLE Projekt.Dim_TimeOfDay ADD
```

```
CONSTRAINT PK_Dim_TimeOfDay PRIMARY KEY (TimeOfDayID);
```

```
-- Dim_WardDemographics IF
```

```
NOT EXISTS
```

```
( SELECT * FROM sys.key_constraints
  WHERE name = 'PK_Dim_WardDemographics' AND parent_object_id =
OBJECT_ID('Projekt.Dim_WardDemographics') )
```

```
ALTER TABLE Projekt.Dim_WardDemographics ADD
```

```
CONSTRAINT PK_Dim_WardDemographics PRIMARY KEY (Ward);
```

```
-- Dim_IUCR
```

```
IF NOT EXISTS
```

```
( SELECT * FROM sys.key_constraints
  WHERE name = 'PK_Dim_IUCR' AND parent_object_id = OBJECT_ID('Projekt.Dim_IUCR')
)
```

```
ALTER TABLE Projekt.Dim_IUCR ADD
```

```
CONSTRAINT PK_Dim_IUCR PRIMARY KEY (IUCR_Code);
```

```
-- Fact_Crime
```

```
IF NOT EXISTS (
```

```
SELECT * FROM sys.key_constraints WHERE
name = 'PK_Fact_Crime' AND parent_object_id = OBJECT_ID('Projekt.Fact_Crime')
)
ALTER TABLE Projekt.Fact_Crime ADD
CONSTRAINT PK_Fact_Crime PRIMARY KEY (CrimeID);

-- FOREIGN KEYS

-- Fact_Crime.DateID ѕ Dim_Date IF NOT
EXISTS ( SELECT *
    FROM sys.foreign_keys WHERE name =
    'FK_Fact_Crime_Date'
)
ALTER TABLE Projekt.Fact_Crime ADD
CONSTRAINT FK_Fact_Crime_Date FOREIGN
KEY (DateID) REFERENCES Projekt.Dim_Date(DateID);

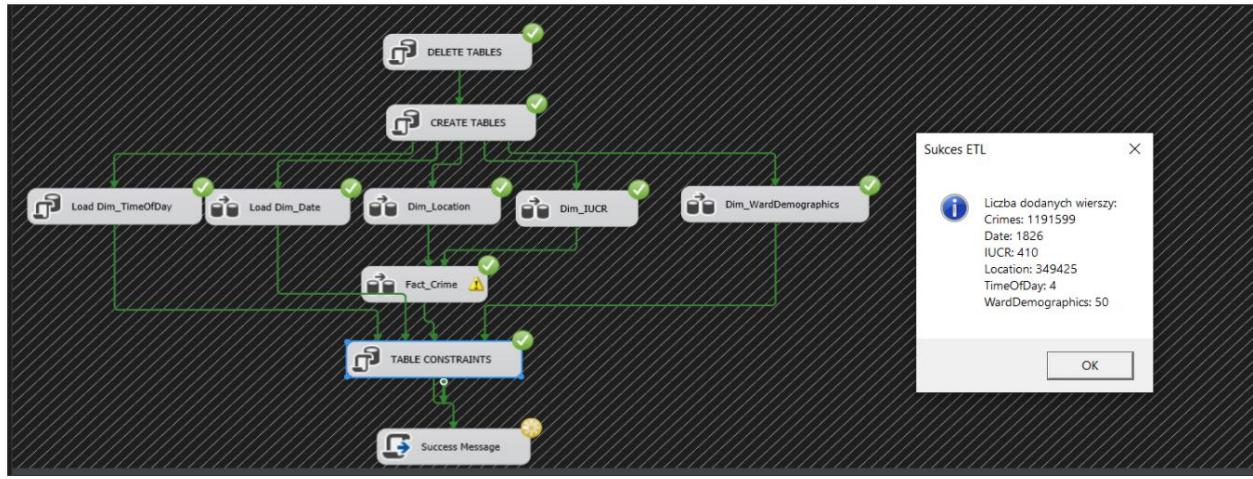
-- Fact_Crime.TimeOfDayID ѕ Dim_TimeOfDay IF NOT
EXISTS ( SELECT *
    FROM sys.foreign_keys WHERE name =
    'FK_Fact_Crime_TimeOfDay'
)
ALTER TABLE Projekt.Fact_Crime ADD
CONSTRAINT FK_Fact_Crime_TimeOfDay FOREIGN
KEY (TimeOfDayID) REFERENCES Projekt.Dim_TimeOfDay(TimeOfDayID);

-- Fact_Crime.LocationID ѕ Dim_Location IF NOT
EXISTS ( SELECT *
    FROM sys.foreign_keys WHERE name =
    'FK_Fact_Crime_Location'
)
ALTER TABLE Projekt.Fact_Crime ADD
CONSTRAINT FK_Fact_Crime_Location FOREIGN
KEY (LocationID) REFERENCES Projekt.Dim_Location(LocationID);

-- Fact_Crime.IUCR ѕ Dim_IUCR IF
NOT EXISTS
( SELECT * FROM sys.foreign_keys
WHERE name = 'FK_Fact_Crime_IUCR'
)
ALTER TABLE Projekt.Fact_Crime ADD
CONSTRAINT FK_Fact_Crime_IUCR
FOREIGN KEY (IUCR) REFERENCES Projekt.Dim_IUCR(IUCR_Code);

-- Dim_Location.Ward ѕ Dim_WardDemographics.Ward IF NOT
EXISTS ( SELECT *
    FROM sys.foreign_keys WHERE name =
    'FK_Location_WardDemographics'
)
ALTER TABLE Projekt.Dim_Location ADD
CONSTRAINT FK_Location_WardDemographics FOREIGN
KEY (Ward) REFERENCES Projekt.Dim_WardDemographics(Ward);
```

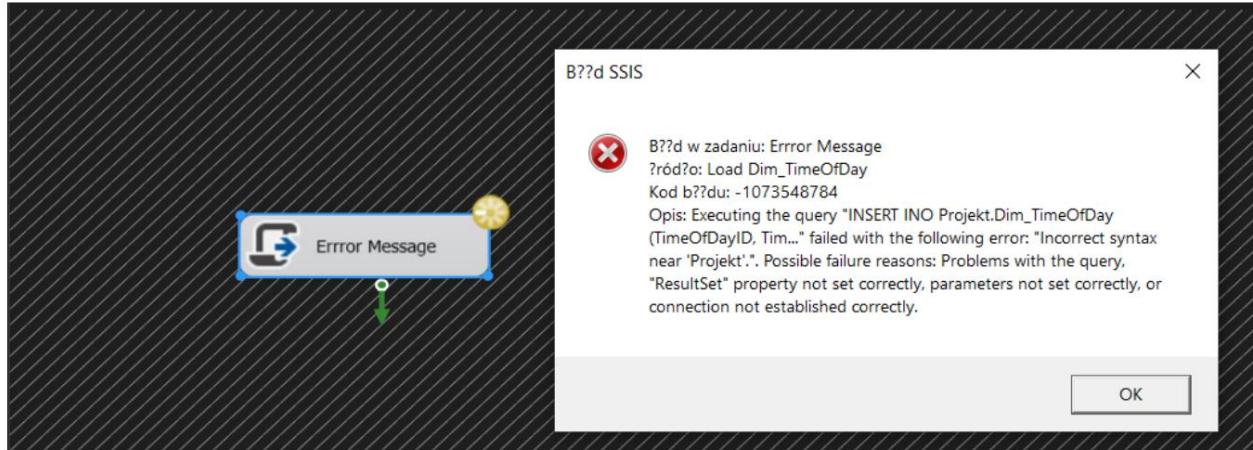
10. Success Message



Purpose: To notify the user about the successful completion of the ETL process and print the number of rows in each table.

Description: C# script that uses SSIS variables corresponding to the number of rows (populated with RowCount).

11. Error Handling



Purpose: To inform the user about an error in the ETL process and print the error code and error description.

Description: A C# script that uses system variables related to Error Handling.

Step 3 – OLAP Cube in SSAS

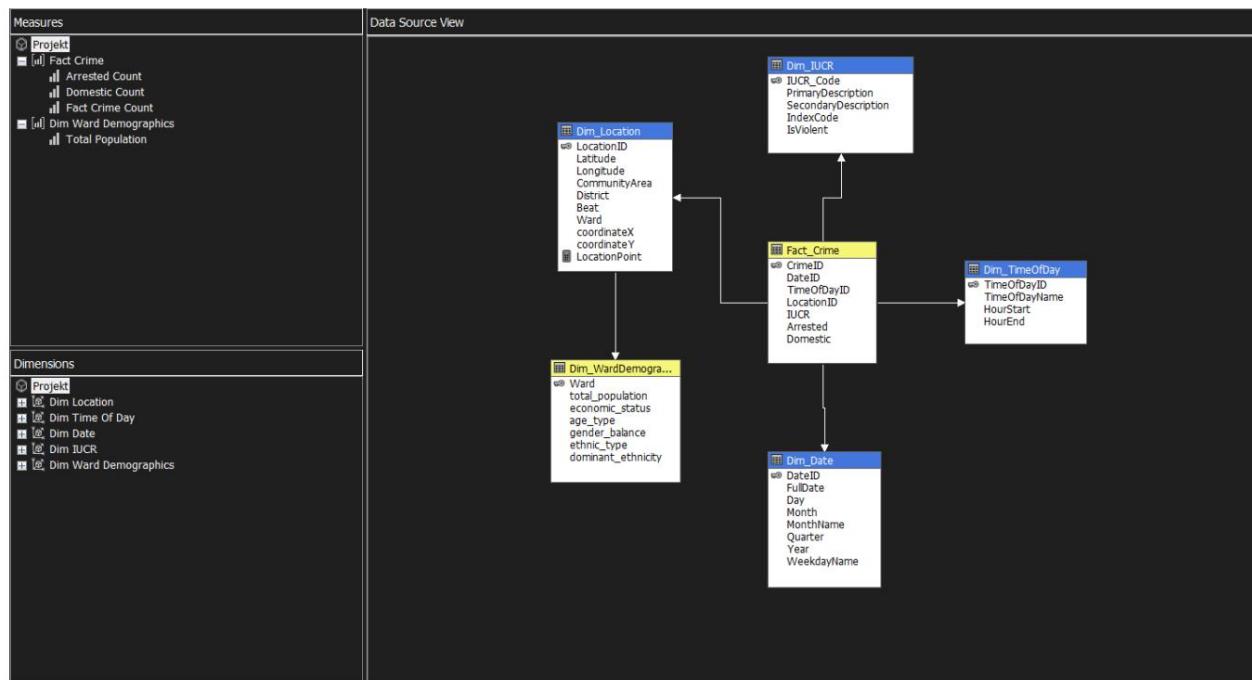
1. Stage Goal Description

The goal of the third stage was to build an OLAP cube in SQL Server Analysis Services (SSAS) that allows for multidimensional analysis of crime in Chicago from 2018 to 2022, taking into account socio-economic data. The cube was built based on data from a previously prepared warehouse, fed by an ETL process implemented in SSIS.

The purpose of the cube is to enable analysis of, among others:

- types of crimes,
- effectiveness of arrests,
- crime in the context of demographic conditions,
- spatial and temporal dependencies.

2. Cube Structure



Fact table:

- Fact_Crime – contains information about individual crime cases.

Measures (facts) - additive:

- Crime Count – number of all crimes (number of records).
- Arrested Count – number of crimes ending with an arrest.
- Domestic Count – number of crimes of a domestic nature.

Calculated Measures:

The screenshot shows the Script Organizer interface. On the left, there's a tree view with items like 'Command', 'CALCULATE', '[Arrest Rate]', '[Domestic Rate]' (selected), and '[Crimes per 1000 Inhabitants]'. The main area has tabs for 'Name' (set to '[Domestic Rate]'), 'Parent Properties' (set 'Parent hierarchy: Measures'), 'Expression' (containing the DAX code), and 'Additional Properties' (with a 'Format string:' dropdown). A status bar at the bottom right shows 'Ln: 5 Ch: 2 SPC CRLF'.

- Arrest Rate – the ratio of Arrested Count to Crime Count
- Domestic Rate – the ratio of Domestic Count to Crime Count
- Crimes per 1000 Inhabitants – a measure calculated as (Crime Count / ward population)
* 1000 (non-additive).

3. Dimensions

Dim_Date:

The screenshot shows the Dimension Editor for the 'Dim_Date' dimension. It has three main panes: 'Attributes' (listing DateID, Date ID, Day, Full Date, Month, Month Name, Quarter, Weekday Name, and Year), 'Hierarchies' (showing a hierarchy structure with levels Year, Quarter, Month Name, and Day, and a note to create a new hierarchy by dragging an attribute), and 'Data Source View' (listing DateID, FullDate, Day, Month, MonthName, Quarter, Year, and WeekdayName).

- Attributes:
 - o DateID - unique date identifier, o FullDate
- full date (format YYYY-MM-DD), o Year, Quarter, Month, Day - numbers defining time units, o MonthName - month name, o WeekdayName - day of the week
(e.g. Monday),
- Hierarchy: Year ÿ Quarter ÿ Month ÿ Day

Dim_TimeOfDay

Attributes	Hierarchies	Data Source View
<input checked="" type="checkbox"/> Dim Time Of Day <ul style="list-style-type: none"> <input type="checkbox"/> Time Of Day ID <input type="checkbox"/> Time Of Day Name 	To create a new hierarchy, drag an attribute here.	 <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Dim_TimeOfDay <input type="checkbox"/> TimeOfDayID <input type="checkbox"/> TimeOfDayName <input type="checkbox"/> HourStart <input type="checkbox"/> HourEnd

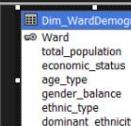
- Attributes:
 - TimeOfDayID - time of day identifier
 - o TimeOfDayName – name of the time of day (e.g. Night, Morning)
 - No hierarchy (flat dimension)

Dim_Location

Attributes	Hierarchies	Data Source View
<input checked="" type="checkbox"/> Dim Location <ul style="list-style-type: none"> <input type="checkbox"/> Beat <input type="checkbox"/> Community Area <input type="checkbox"/> District <input type="checkbox"/> Location ID <input type="checkbox"/> Location Point <input type="checkbox"/> Ward 	 <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Police Geography <ul style="list-style-type: none"> = District = Beat <new level> To create a new hierarchy, drag an attribute here.	 <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Dim_Location <input type="checkbox"/> LocationID <input type="checkbox"/> Latitude <input type="checkbox"/> Longitude <input type="checkbox"/> CommunityArea <input type="checkbox"/> District <input type="checkbox"/> Beat <input type="checkbox"/> Ward <input type="checkbox"/> coordinateX <input type="checkbox"/> coordinateY <input type="checkbox"/> LocationPoint

- Attributes:
 - o Ward, District, Beat, CommunityArea – administrative division units cities
 - o RoundedLatitude, RoundedLongitude – coordinates rounded to 4 decimal places – used for aggregation on maps
- Hierarchies:
 - o **Police Geography:** District > Beat

Dim_WardDemographics

Attributes	Hierarchies	Data Source View
<input checked="" type="checkbox"/> Dim Ward Demographics <ul style="list-style-type: none"> <input type="checkbox"/> Age Type <input type="checkbox"/> Dominant Ethnicity <input type="checkbox"/> Economic Status <input type="checkbox"/> Ethnic Type <input type="checkbox"/> Gender Balance <input type="checkbox"/> Total Population <input type="checkbox"/> Ward 	To create a new hierarchy, drag an attribute here.	 <ul style="list-style-type: none"> <input checked="" type="checkbox"/> Dim_WardDemographic... <input type="checkbox"/> Ward <input type="checkbox"/> total_population <input type="checkbox"/> economic_status <input type="checkbox"/> age_type <input type="checkbox"/> gender_balance <input type="checkbox"/> ethnic_type <input type="checkbox"/> dominant_ethnicity

- Attributes:
 - o economic_status – ward classification into poor / medium / rich
 - o age_type – dominant age group (e.g. young-majority)
 - about gender_balance – gender balance

- o ethnic_type – ethnic type (mixed / dominant)
- o dominant_ethnicity – dominant ethnic group
- o total_population – the number of ward inhabitants
- No hierarchy – descriptive data for correlational analysis

Dim_IUCR

Attributes	Hierarchies	Data Source View
<input checked="" type="checkbox"/> Dim_IUCR <ul style="list-style-type: none"> <input type="checkbox"/> Index Code <input type="checkbox"/> Is Violent <input type="checkbox"/> IUCR Code <input type="checkbox"/> Primary Description <input type="checkbox"/> Secondary Description 	Hierarchy <ul style="list-style-type: none"> Primary Description Secondary Description <new level> <p>To create a new hierarchy, drag an attribute here.</p>	Dim_IUCR <ul style="list-style-type: none"> IUCR_Code PrimaryDescription SecondaryDescription IndexCode IsViolent

- Attributes:
 - o PrimaryDescription, SecondaryDescription – basic and detailed classification of crime
 - o IndexCode – reporting code
 - o IsViolent – Violent Crime Flag
 - o CrimeCategory – parent categorization (e.g. Property Crime, Violent Crime)
 - o SeverityWeight – the severity of the crime, prepared based on PrimaryDescription, in order to calculate weighted indicators (e.g. KPI)
- Hierarchy: CrimeCategory -> PrimaryDescription ÿ SecondaryDescription

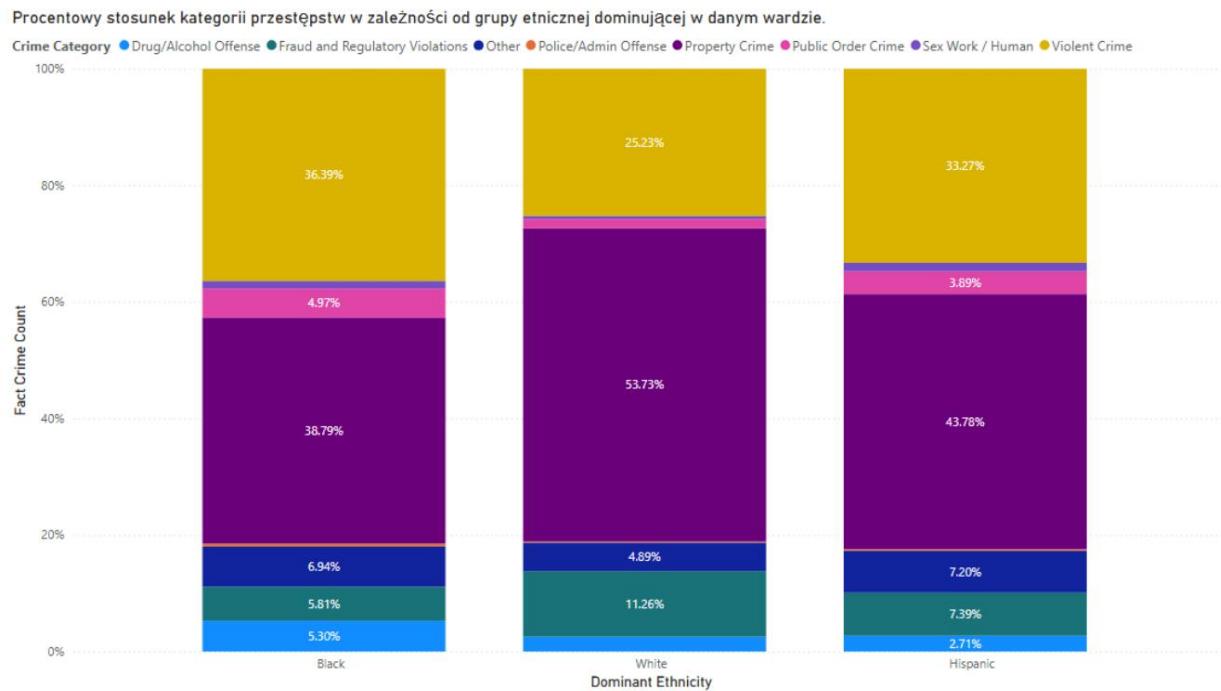
Dim_CrimeFlags (degenerate dimension):

Hierarchies						Data Source View
Attributes						
	Name	Usage	Type	Key Column	Name Column	
<input type="checkbox"/>	Arrested	Regular	Regular	Integer	Same as key	
<input type="checkbox"/>	Crime ID	Key	Regular	Integer	Same as key	
<input type="checkbox"/>	Domestic	Regular	Regular	Integer	Same as key	
	To create a new attribut...					

- Attributes:
 - o Arrested – arrest flag (True/False)
 - o Domestic – Domestic crime flag (True/False)
- Notes: This dimension is a so-called **degenerate dimension** – its attributes come directly from the fact table, but are separated for analytical purposes and model readability.

4. Analyses – Stage 3 (implementation of point 7 of Stage 1)

Analysis 1: Distribution of crime categories by dominant ethnic group



Conclusions:

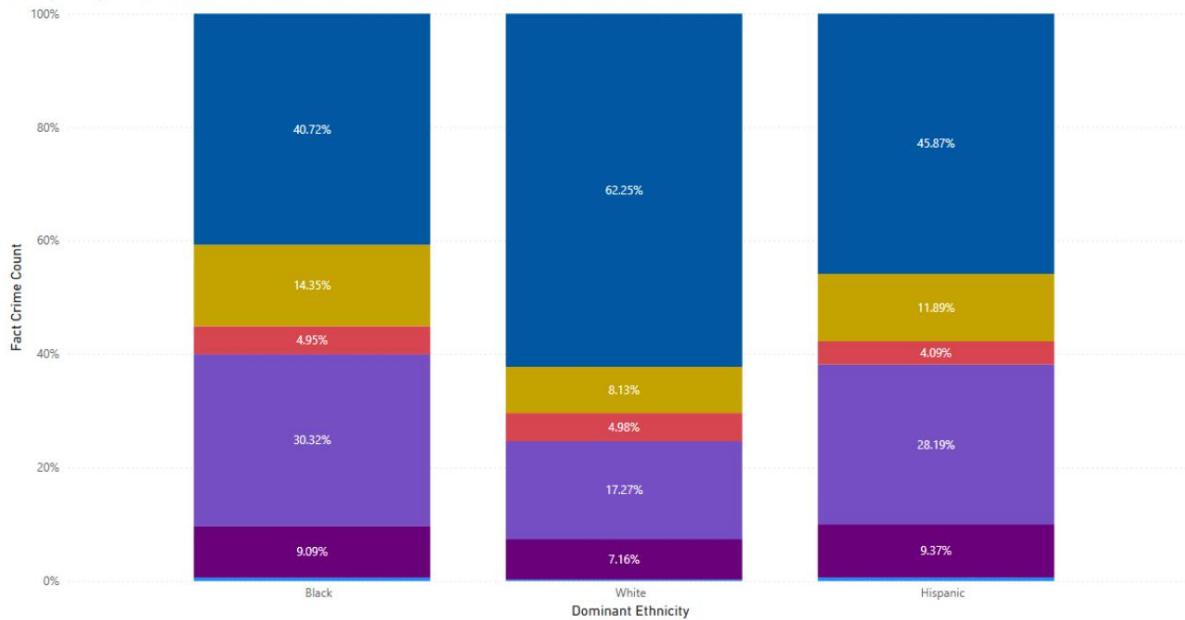
- The structure of crimes in wards with different dominant ethnic groups is generally similar — in all cases, **property crimes** (burglary, theft, property damage) and **violent crimes** (assaults, beatings, murders) predominate.
- In wards with a majority **White (non-Hispanic) population**, the highest percentage of **property crimes** is particularly noticeable , accounting for **53.73%** of all crimes. This is significantly higher than in wards with a majority **Black or African American** or Hispanic or Latino population .
- Wards with a dominant **White (non-Hispanic)** population also have significantly higher rates of **Fraud and Regulatory Violations** and lower rates of violent crime than wards with a dominant Black or Latino population.

Let us therefore analyze the distribution of crime types in the two most popular categories.

Detailed breakdown in the "Property Crime" category:

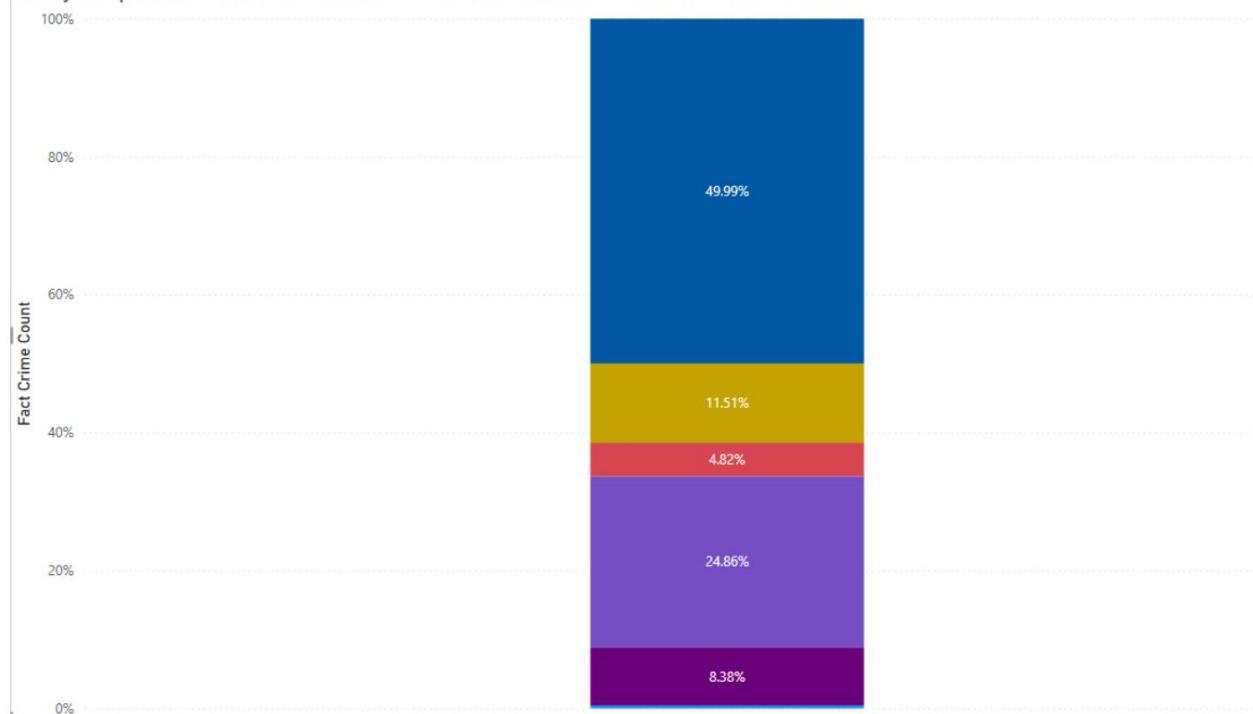
Procentowy stosunek typu przestępstwa własnościowego w zależności od grupy etnicznej dominującej w danym wardzie.

Primary Description ● ARSON ● BURGLARY ● CRIMINAL DAMAGE ● CRIMINAL TRESPASS ● MOTOR VEHICLE THEFT ● THEFT



Procentowy stosunek typu przestępstwa własnościowego.

Primary Description ● ARSON ● BURGLARY ● CRIMINAL DAMAGE ● CRIMINAL TRESPASS ● MOTOR VEHICLE THEFT ● THEFT

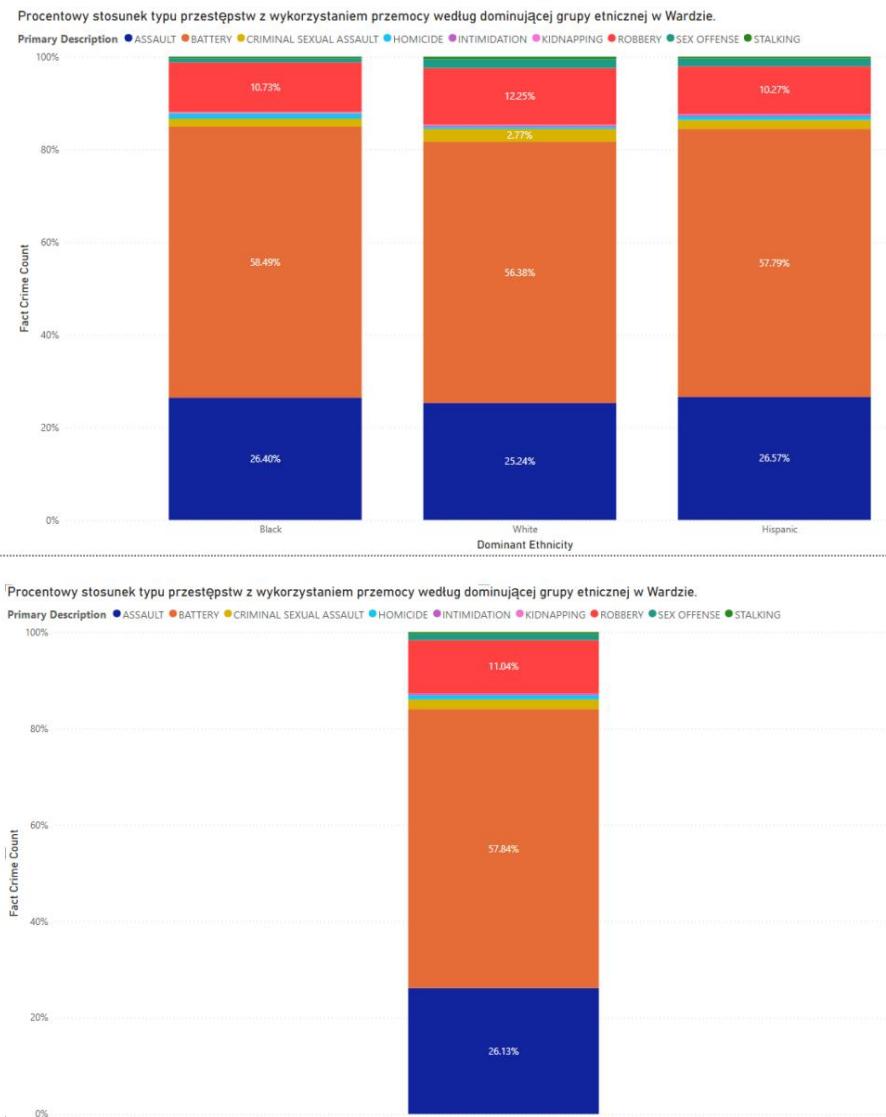


- The distribution of property crime types in wards with dominant **Black or African American** and **Hispanic or Latino** populations is very similar — the structure of these crimes does not differ significantly between these groups.

- In wards with a dominant **White (non-Hispanic)** population , however, **a much greater dominance of Theft**, with a lower share of **Criminal Damage** and **Motor Vehicle Theft** compared to other wards.
- Detailed breakdown within the “Property Crime” category:
 - **Theft: 50.00%**
 - **Property damage (Criminal Damage): 24.86%**
 - **Motor Vehicle Theft: 11.51%**

This may indicate differences in the availability and security of property depending on the dominant social group or the specific nature of local property crimes.

Detailed Analysis: Violent Crime

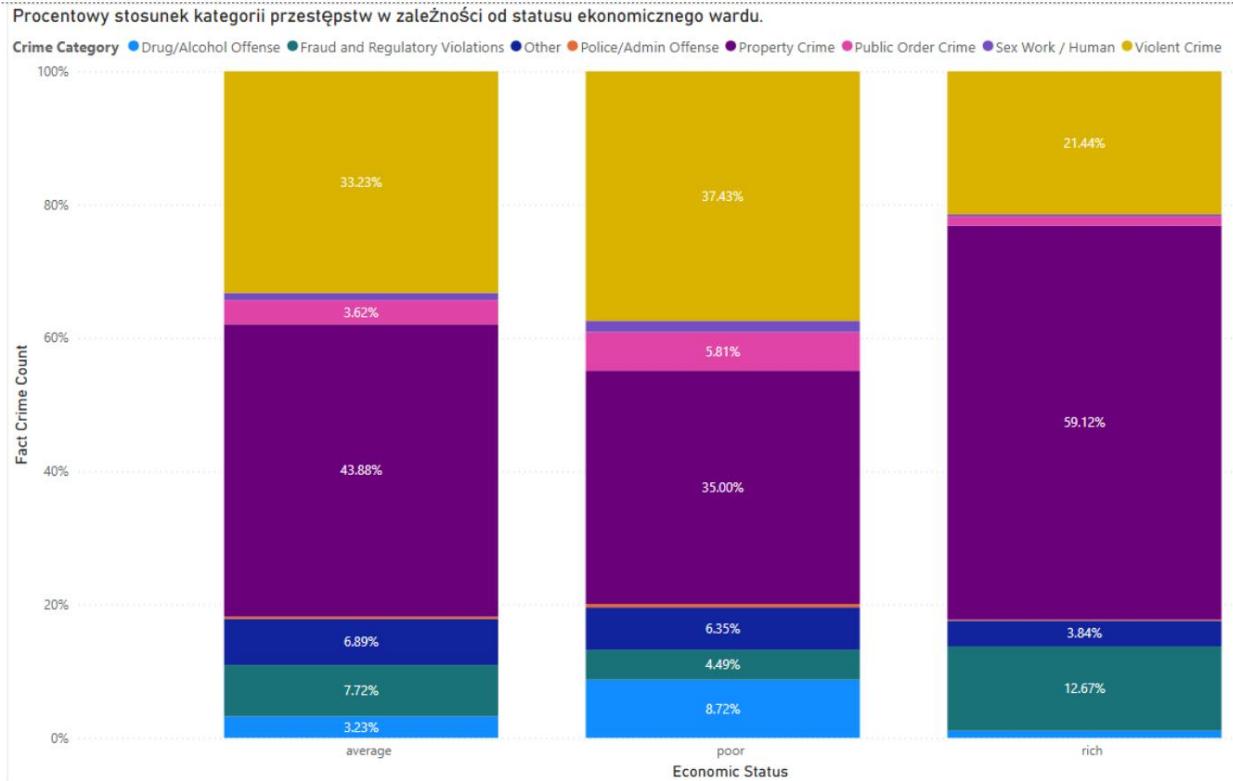


- Distribution of types of violent crimes in wards with dominant groups
Black or African American, Hispanic or Latino, and White (non-Hispanic) are very similar — the most common are:
 - **Beatings (Battery): 57.84%**
 - **Assault: 26.13%**
 - **Robbery: 11.04%**
- It is worth noting, however, that **in wards with a dominant White population (non-Hispanic)** the share of sexual assaults (Criminal Sexual Assault) is significantly higher than in the other:

- o **White (non-Hispanic): 2.77%**
- o **Black or African American: 1.71%**
- Hispanic or Latino: **2.08%**

These differences may be due to both local circumstances and the way sexual crimes are reported and classified in different communities.

Analysis 2: Types of Crimes vs. Ward Economic Status

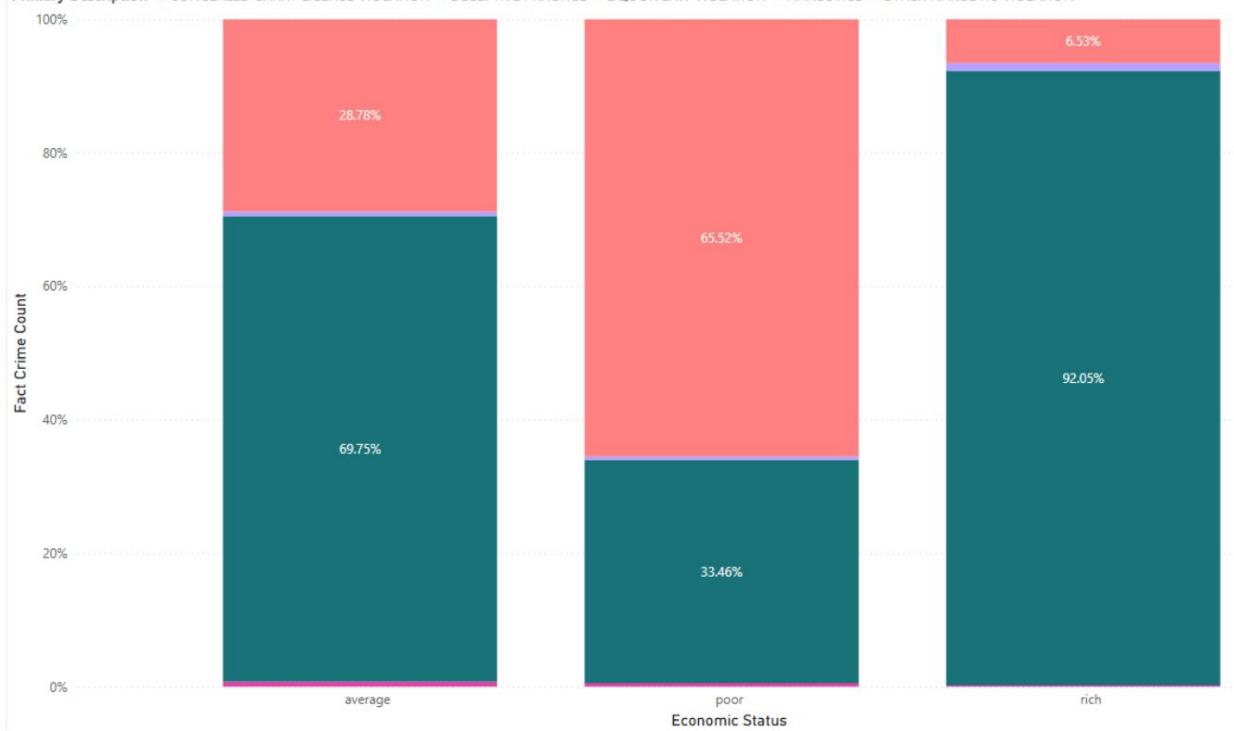


- The structure of crimes differs significantly depending on the wealth level of the district.
 - In **poor** wards, a **higher share of violent crimes** dominates (Violent Crime – 37.43%) compared to rich wards (21.44%) and medium-wealth wards (33.23%).
- In **wealthy** wards , the category of **property crime** clearly dominates , accounting for **59.12%** of all crimes – significantly more than in poor wards (35.00%) and middle-income wards (43.88%). This may suggest that in wealthier districts opportunistic crime (thefts, burglaries) dominates, not aggressive crime.
- **Public Order Crime** and **Drug/Alcohol Offense** are **most prevalent in poor wards**, which may indicate more frequent preventive policing or greater exposure of residents in these areas to such interventions.
- **Fraud and Regulatory Violations** occur much more often in middle-income (7.72%) and wealthy (12.67%) wards than in poor wards (4.49%).

Let's examine Fraud and Regulatory Violations and Drug/Alcohol Offense because of the apparent difference between poor/middle-income and wealthy wards.

Procentowy stosunek typu przestępstw formalnych oraz narkotykowych w zależności od statusu ekonomicznego wardu.

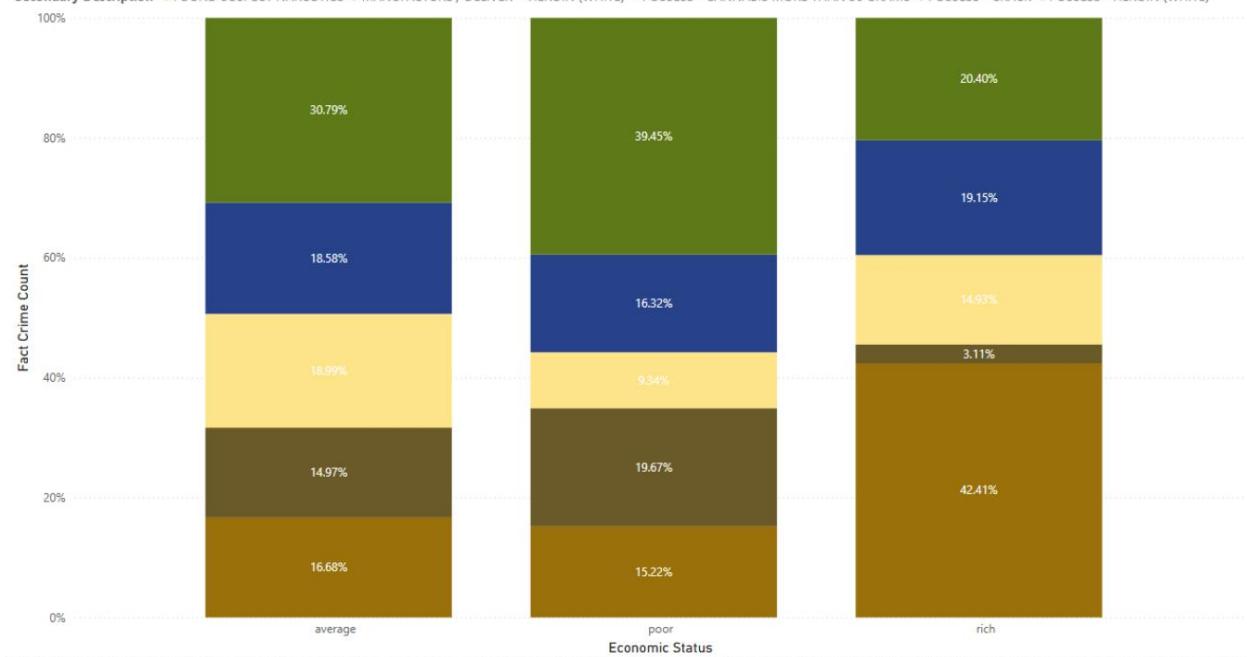
Primary Description ● CONCEALED CARRY LICENSE VIOLATION ● DECEPTIVE PRACTICE ● LIQUOR LAW VIOLATION ● NARCOTICS ● OTHER NARCOTIC VIOLATION



We see that drug crimes dominate in poor wards, while fraud (financial and testimony) dominates in middle- and wealthy wards.

Procentowy rozkład top 5 przestępstw narkotykowych według typu ekonomicznego Wardu.

Secondary Description ● FOUND SUSPECT NARCOTICS ● MANUFACTURE / DELIVER - HEROIN (WHITE) ● POSSESS - CANNABIS MORE THAN 30 GRAMS ● POSSESS - CRACK ● POSSESS - HEROIN (WHITE)



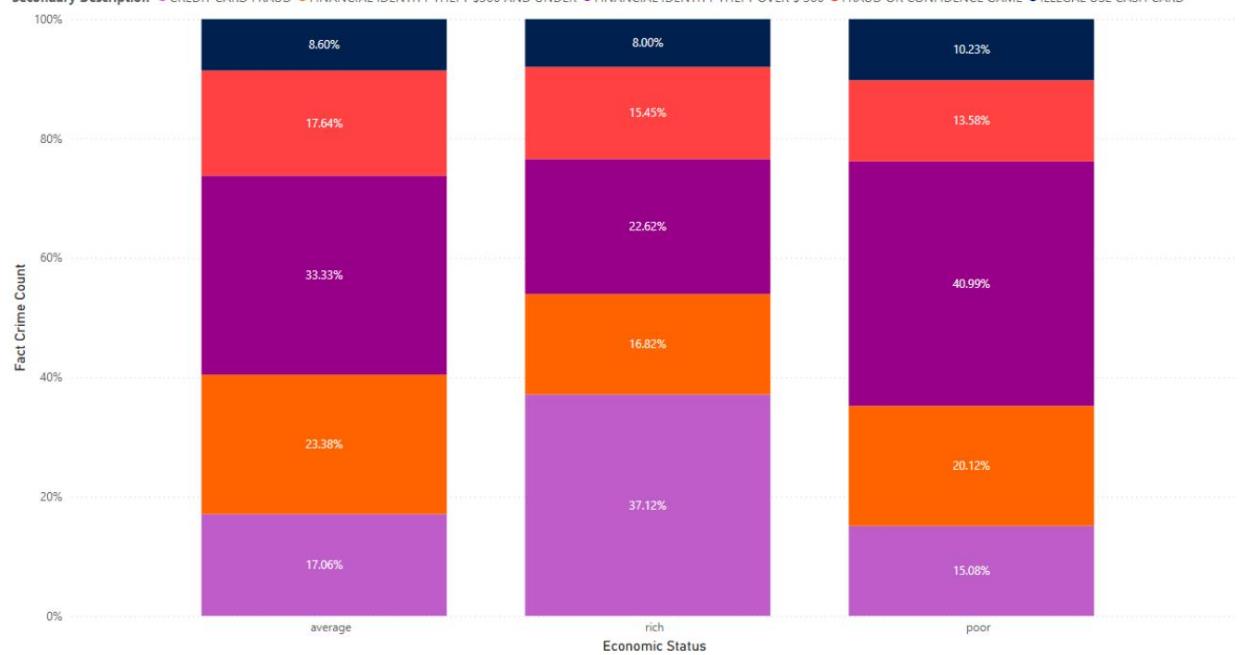
The five most common drug-related crimes, based on the ward's economic status, included the following categories:

- **Possession of heroin** is the most common type of drug offence in **wards of low economic status**.
- **Crack Possession** – A regular occurrence in every ward group.
- **Possession of more than 30 grams of marijuana** – an important category, especially in wards middle-income and rich.
- **Heroin production and distribution** – occurs more frequently in **poor and middle-income wards**, which may indicate the presence of organized drug crime in these areas.
- **Possession of suspicious substances** – most frequently recorded in **wealthy wards**, where it is responsible for as much as **42% of drug crimes**.

The high proportion of “possession of suspected substances” offences in wealthier wards may suggest **a different way of classifying offences or a less severe approach by law enforcement**. It may also indicate **differences in policing practices or potential bias towards more affluent communities** – which would be worth further exploring in the context of **social and systemic justice**.
equality.

Procentowy rozkład top 5 oszustw i przestępstw na tle legalnym według typu ekonomicznego Wardu.

Secondary Description ● CREDIT CARD FRAUD ● FINANCIAL IDENTITY THEFT \$300 AND UNDER ● FINANCIAL IDENTITY THEFT OVER \$ 300 ● FRAUD OR CONFIDENCE GAME ● ILLEGAL USE CASH CARD



The most common crimes in the category of **fraud and formal and legal violations** vary depending on the economic status of the wards:

- **Financial Identity Theft over \$300** accounts for the highest percentage of this category in low- and middle-income wards:
 - **Poor wards:** as much as **40.99%**
 - **Average wards:** **33.33%**
 - It occurs less frequently in wealthier wards
- **Financial identity theft up to \$300 (\$300 and under)** is also present, but not clearly dominant in any economic class
- **Credit Card Fraud** is **much more common in wards wealthy:**
 - **Rich wards:** **37.12%**
 - For comparison: only **15.08%** in poor wards and **17.06%** in middle-income
- **Fraud related to games of chance and the use of payment cards** (*Confidence Game, Illegal Use of Cash Card*) occur in **similar proportions** regardless of the economy class of the ward

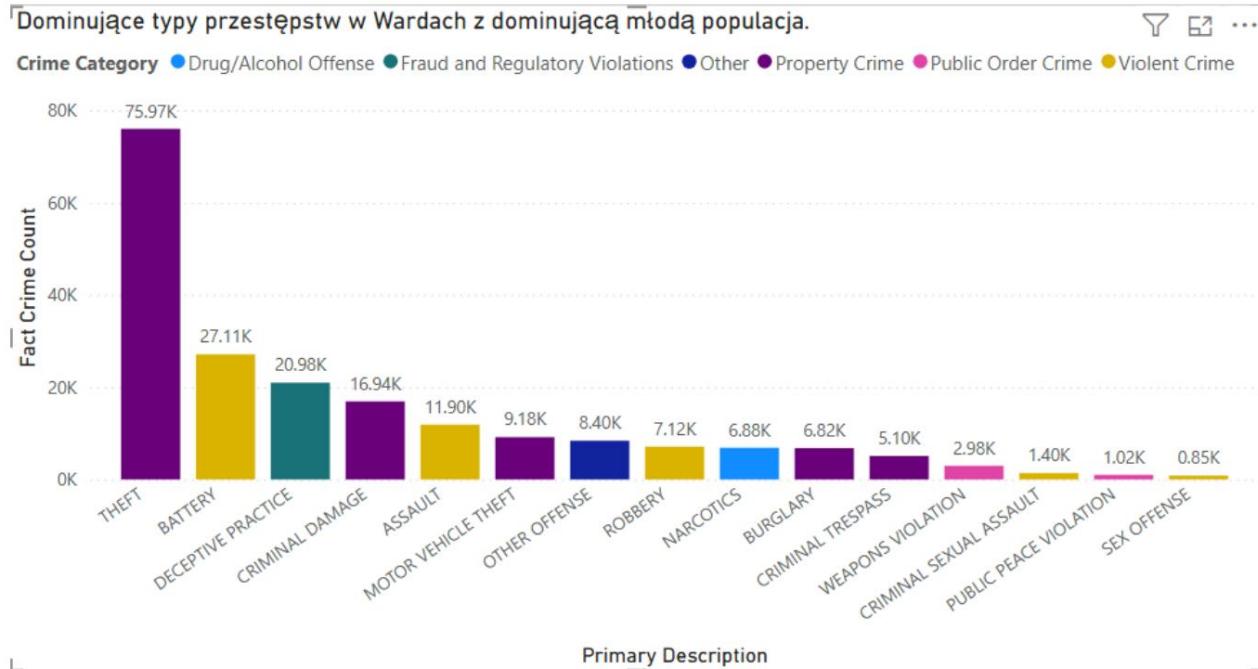
Conclusions: Type of crime and economic status of the ward

Data analysis indicates **significant relationships between the economic class of wards and the structure of crimes committed:**

- **In the poorer wards**, classical, more direct methods predominate crimes such as:
 - **Possession of heroin**
 - **Financial Identity Theft Over \$300**
 - **Drug production and distribution**
- **In wealthy wards**, the following are more common:
 - **Electronic and credit card fraud**
 - **Possession of suspicious substances**, which may be the result of less clear law enforcement or a different classification of offences
- **Middle-income wards** are largely concerned with crimes of a mixed nature, including possession of marijuana in larger quantities or theft of personal data on a smaller scale

These findings suggest that **the crime profile reflects local socio-economic conditions**, and perhaps also **differences in how crimes are classified and prosecuted** across neighborhoods. These data provide a starting point for further research into the equity of the criminal justice system and the hidden mechanisms by which particular social groups are privileged or marginalized.

Analysis 3: Crime in Wards with Young Populations



Analysis of data for wards where **the dominant demographic group is young people (young-majority)** reveals a characteristic crime profile:

- **The most common type of crime is theft**, which occurred **75.97 thousand times**. This means that **property crimes dominate** in this group of wards.
- The next places are occupied by violent crimes:
 - **Beatings ("Battery") – 27.11 thousand cases**
 - **Assault – 11.90 thousand cases**

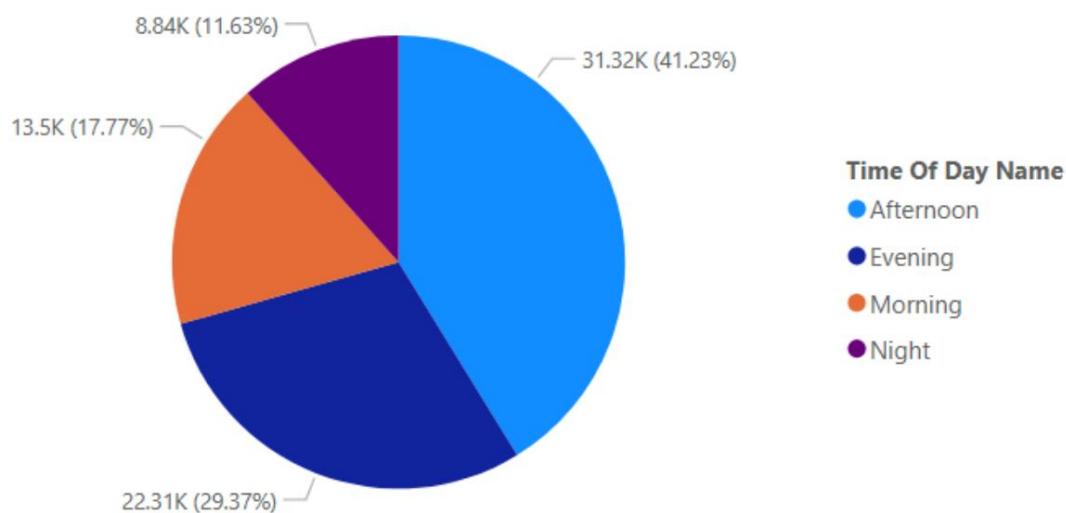
This suggests that the young population is **more vulnerable to both property loss and physical violence**.

- **Offences relating to fraud and regulatory breaches** (e.g. (Deceptive Practice) occurred **20.98 thousand times**, which indicates **a significant presence of economic crime** among young residents - perhaps related to the greater share of digital transactions and consumerist lifestyle.
- **Drug-related crimes ("Narcotics")**, although less numerous (**6.88 thousand**), are still constitute **a noticeable element of local crime**.

- The high number of non-robbery thefts may be related to the activities of pickpockets and petty thieves, especially in the context of mass events, nightlife and parties that are popular among the city's younger residents.

In order to verify this hypothesis, **an additional analysis of the dependence of theft on the time of day will be conducted**, which will allow for assessing whether increased criminal activity occurs at night.

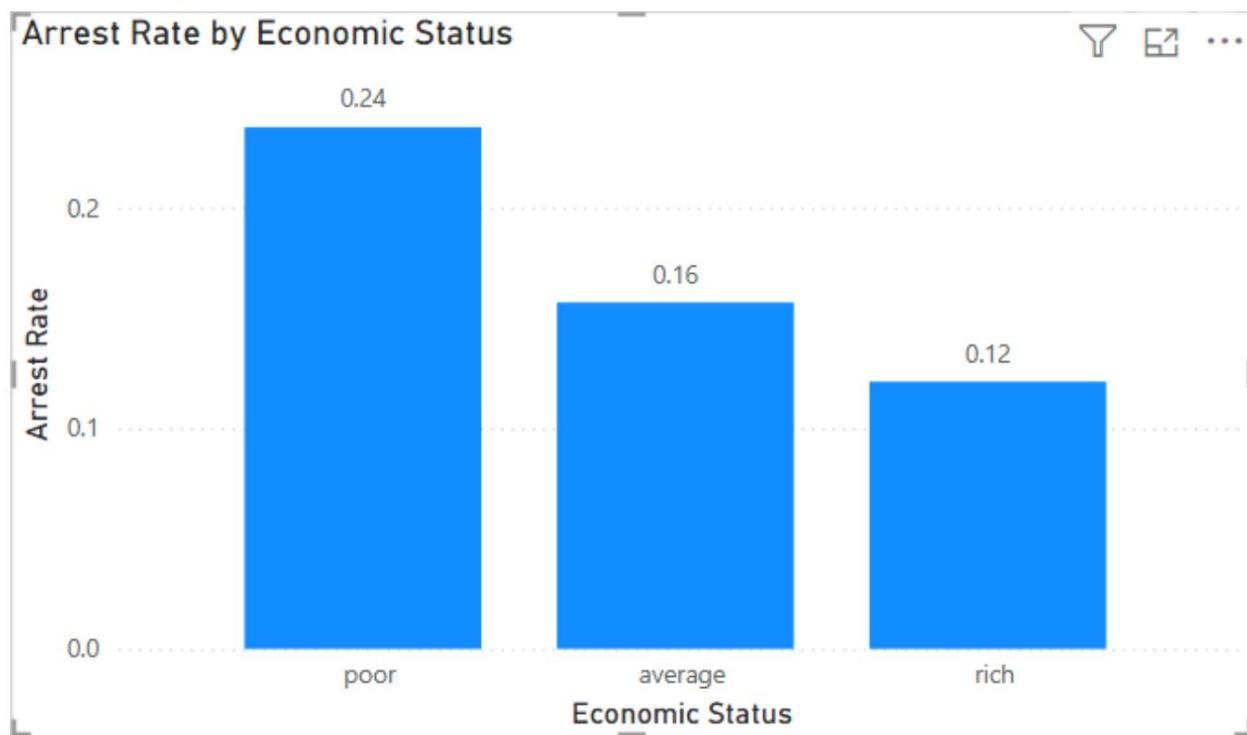
Rozkład czasu dnia kradzieży w Wardach z dominującym młodym społeczeństwem.



The analysis of the distribution of **theft**- type crimes by time of day (TimeOfDayName) **does not clearly confirm the hypothesis** that thefts are most frequent in the evening or at night — i.e. during hours typically associated with events and social activities of young people.

- Most thefts were recorded **in the afternoon (Afternoon)** – 31 323 cases, which is the largest share in this analysis.
- In second place are **the evening hours (Evening)** – 22,312 cases.
- **Morning: 13,497 thefts**
- **At night (Night)** the lowest number of thefts was recorded – 8,836 cases

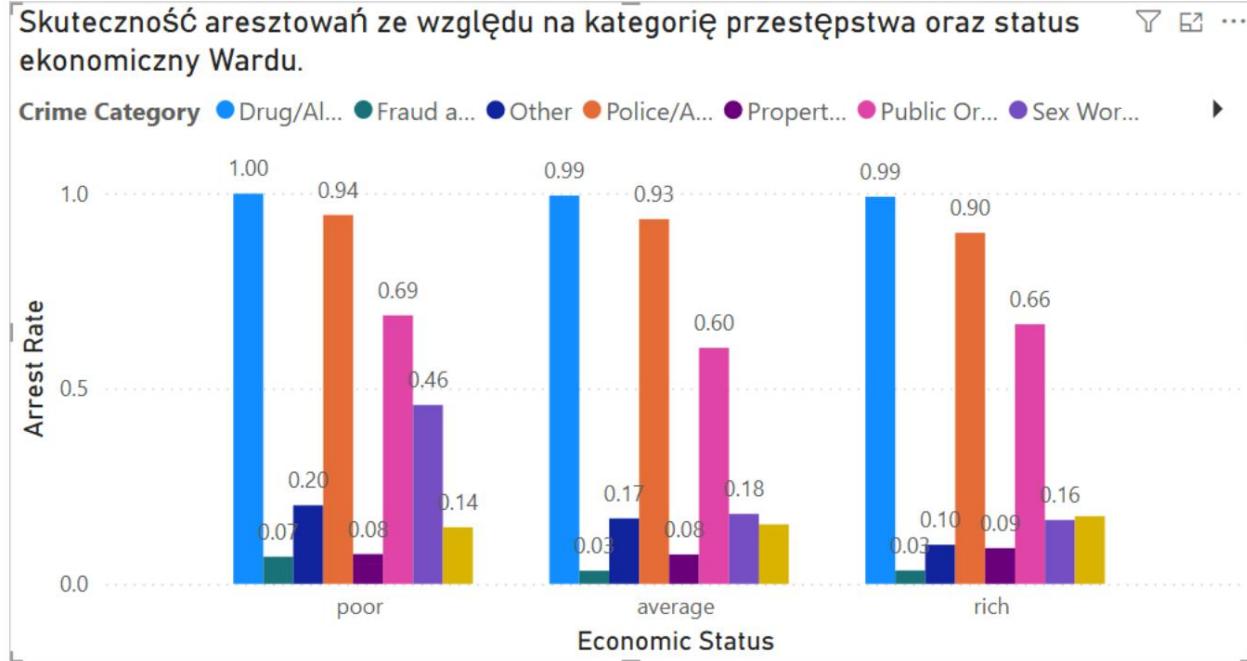
Analysis 4: Arrest Effectiveness by Economic Status



Analysis of Arrest Rate by **Wards' Economic Status** indicates a clear relationship:

- In poor wards: the average arrest rate is 24%
- In average wards: 16%
- In the rich wards: only 12%

These data suggest that **the higher the economic status of a ward, the lower the overall arrest rate**. However, this may be partly due to **differences in the type of crimes committed**.



After extending the analysis with an additional dimension – **the Crime Category** – we notice that:

- In most categories, arrest rates do not differ significantly across ward classes. This suggests that overall differences are more likely to be a result of the distribution of crime types rather than differences in police actions.
- The exception are crimes from the “Sex Work / Human Crimes” category, where a strong discrepancy is visible (this difference will be analyzed later in the report):
 - about Ward poor: 46%
 - Ward average : 18%
 - about Ward rich: 16%
- Crime categories with the highest arrest rates:
 - o Drug/Alcohol Offense: 99–100%
 - ÿ Understandable, because these are crimes that are directly revealed by officers during arrest
 - o Police/Admin Offense: 90–94%
 - ÿ Also resulting from direct contact with the police
 - o Public Order Crime: 60-69%

ŷ Probably also revealed directly (e.g. disturbing public order)

- **Categories of crimes with the lowest arrest rates:**

- **Violent Crime: 8–9%**

ŷ Possible difficulties in identifying the perpetrators, lack of witnesses or collecting evidence

- **Fraud and Regulatory Violations: 3–7%**

ŷ Some of these crimes do not always require arrest (e.g. administrative abuse), some may remain undetected

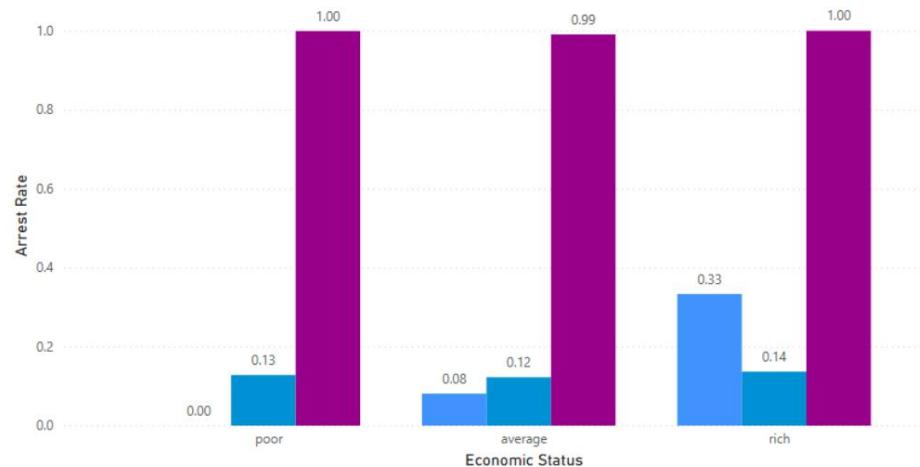
- **Property Crime: 14–18%**

ŷ As in the case of violence – it is difficult to detect perpetrators, especially in the case of thefts without witnesses

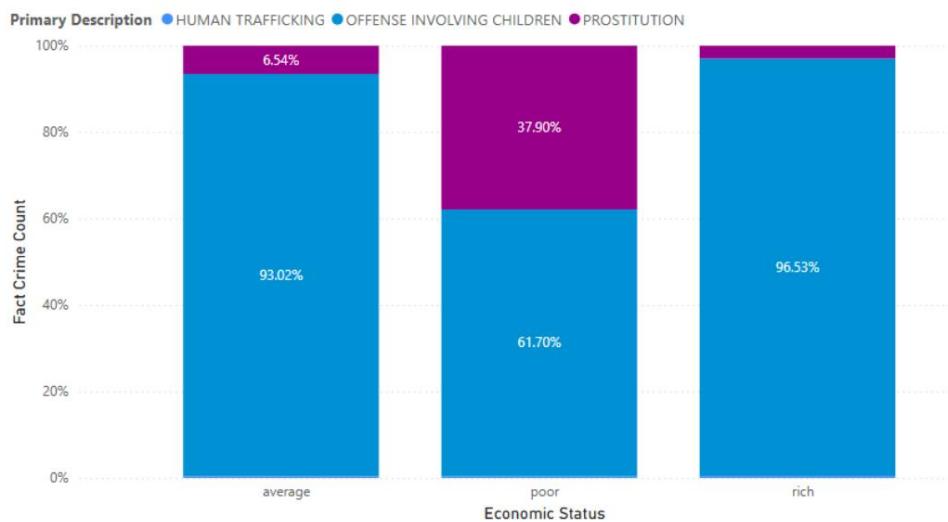
The differences in arrest effectiveness between wards of different economic status are mainly due to the different structure of crime types, and not to the approach of law enforcement agencies themselves. The highest effectiveness concerns crimes in which the perpetrators are **identified directly** by the police, while in the case of property crimes, violence or fraud – **identifying the perpetrators is significantly more difficult**. The significant difference in the category "Sex Work / Human Crimes" may indicate **differences in law enforcement or the purpose of operational activities depending on the ward**, which requires further analysis.

Skuteczność aredzstowań ze względu na typ przestępstwa Sex / Human Crime i status ekonomiczny Wardu.

Primary Description ● HUMAN TRAFFICKING ● OFFENSE INVOLVING CHILDREN ● PROSTITUTION



Stosunek liczb przestępstw ze względu na typ przestępstwa Sex / Human Crime i status ekonomiczny Wardu.



Detailed analysis of crime subcategories within the **Sex Work/Human Crimes** group allows for a better understanding of differences in arrest rates across wards of different economic status.

Effectiveness of arrests by type of crime:

- **Prostitution** – the effectiveness of arrests is very high and practically identical in all wards: **99–100%**
 - o This is most likely due to the direct arrest of the perpetrators place.
- **Offense Involving Children** – 12–14% effectiveness , regardless of class ward
 - o Difficulties in detection, sensitive nature of cases, possible lack of witnesses
- **Human Trafficking** – clear differences:
 - about 0 cases in poor wards
 - 8% of arrests in middle-income wards
 - about 33% of arrests in wealthy wards
 - ŷ This may suggest differences in the approach of the services, unequal effectiveness of investigations, as well as the possibility of institutional factors, such as **underinvestment in services in poorer areas** or potential **socio-political pressure** in wealthier districts. Or simply **too few cases to draw meaningful statistics**.

Structure of crimes:

• **Ward poor:**

about 0 cases of Human Trafficking

37.9 % of crimes are related to prostitution

61.7 % – crimes involving children

• **Ward average:**

about 37 cases of Human Trafficking

by 93.02% – Offense Involving Children

by 6.54% – Prostitution

• **Ward rich:**

about 3 cases of Human Trafficking

by 96.53% – Offense Involving Children

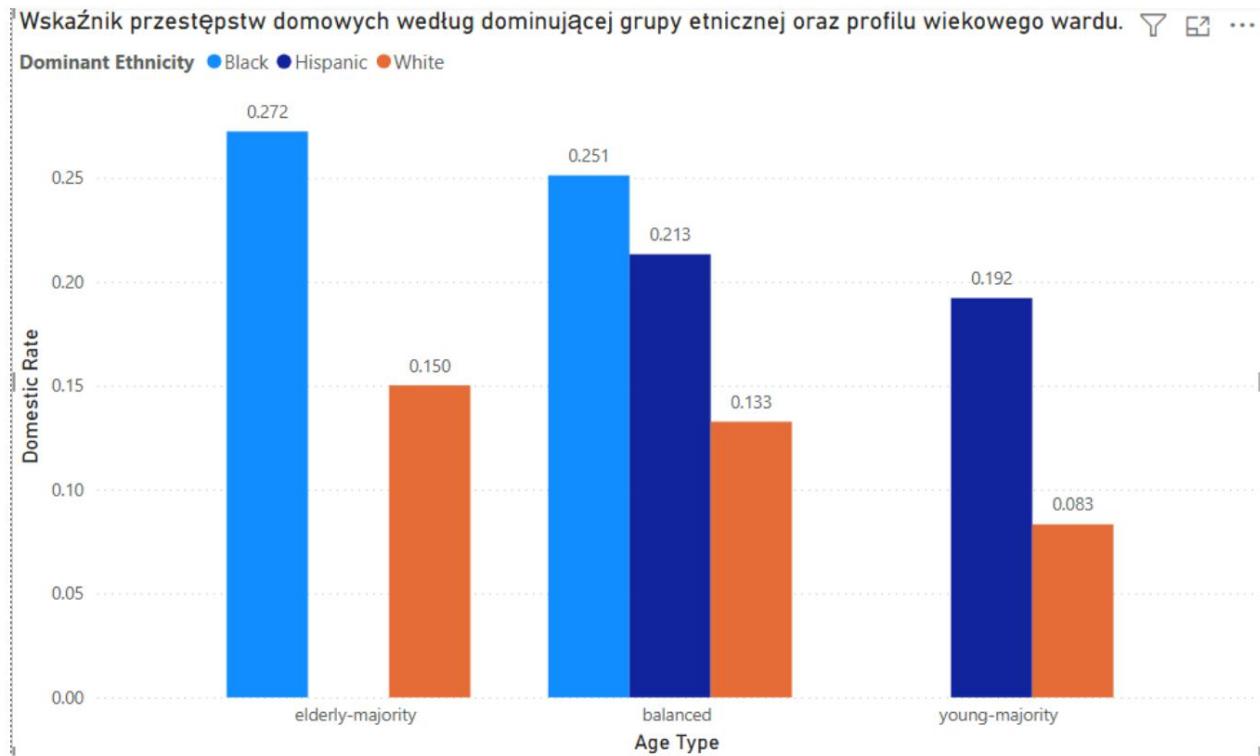
by 2.97% – Prostitution

The differences in the effectiveness of arrests in this category **are not only due to the approach of the services**, but also to **the different crime structures in individual wards**.

Prostitution-related crimes are more common in poor wards, while cases of human trafficking occur only in middle-income and wealthy wards – although very rarely.

Low effectiveness in the most serious cases (e.g. Human Trafficking) may indicate a need for better funding of investigations, development of inter-agency procedures or reveal systemic gaps. High effectiveness of arrests for prostitution contrasts with low effectiveness in more complex and socially significant cases, which may require further reflection.

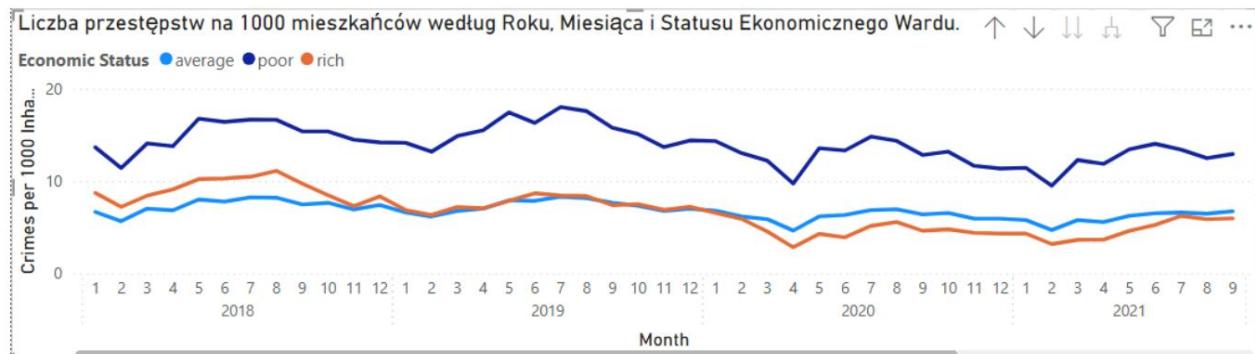
Analysis 5: Domestic Crime Rates and Age and Ethnic Profile



Analysis of the DomesticCrimeRate against the dominant **age** and **ethnic** profile of the wards reveals clear differences in the scale of this phenomenon:

- **The highest rate of domestic crime** (27.2%) was recorded in wards with **an older-majority population and a predominantly Black community**.
- **High levels** (25.1%) are also found in **age-balanced** wards (**balanced**) and **black people**.
- **Hispanic** Wards – both balanced and with a young population – have moderate rates (21.3% and 19.2%), but still higher than white wards.
- **The lowest rates** are found in wards with **a dominant white population (White)** – both in the older (15%) and younger (8.3%) groups.

Analysis 6: Crime Trends by Ward Economic Status



1. Wards with a low economic status (poor) have by far the highest crime rates

- Throughout the 2018–2022 period, wards designated as **poor** maintain **twice or up to three times higher crime rates** than richer wards.
- Example: in July 2019, there were **18.0** crimes per 1,000 inhabitants in poor wards, compared to **8.4** in rich wards.

2. Rich wards have the lowest rate – and consistently so

- In no single month did the rate exceed **11 crimes per 1,000 inhabitants**, and after 2020 it rarely exceeded 7.
- At the peak of the pandemic (April 2020), it even dropped to **2.84**, which is one of the lowest levels in the entire data set.

3. Average wards have moderate crime rates

- The crime rate usually ranges between **6 and 8**.
- Changes are seasonal, but less extreme than in the poor wards.

4. Seasonality visible in all groups, but at different levels

- In the summer months (May–August), rates increase, especially in poorer wards (e.g. **14–18**), which may be related to social activity, weather and more interpersonal interactions.
- During the winter months, rates drop – especially in the rich wards, where the difference between winter and summer can be even **more than 3 percentage points**.

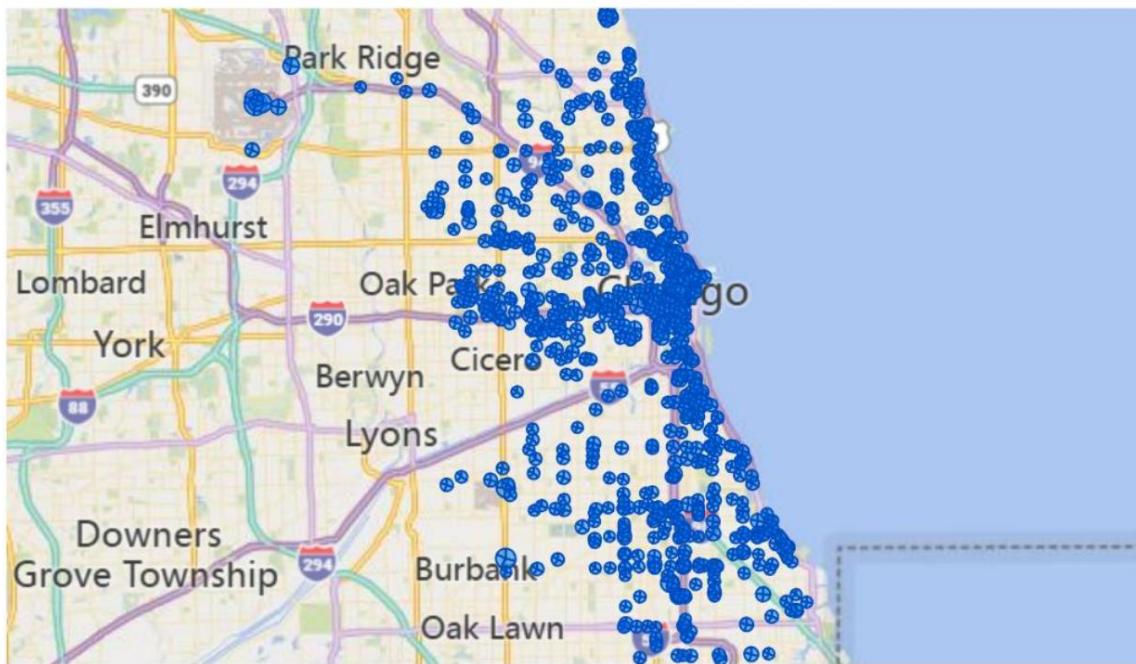
5. The COVID-19 (2020) pandemic has affected all groups, but the richest wards have been the most affected

- The largest decrease in crime per 1,000 inhabitants was observed in the wards of rich – in April 2020 down to **2.83**, compared to ~5–7 in other years.
- Poor wards also recorded declines, but their level was still higher than any result in rich wards.

Analysis 7: Crime rates by day of week and location.

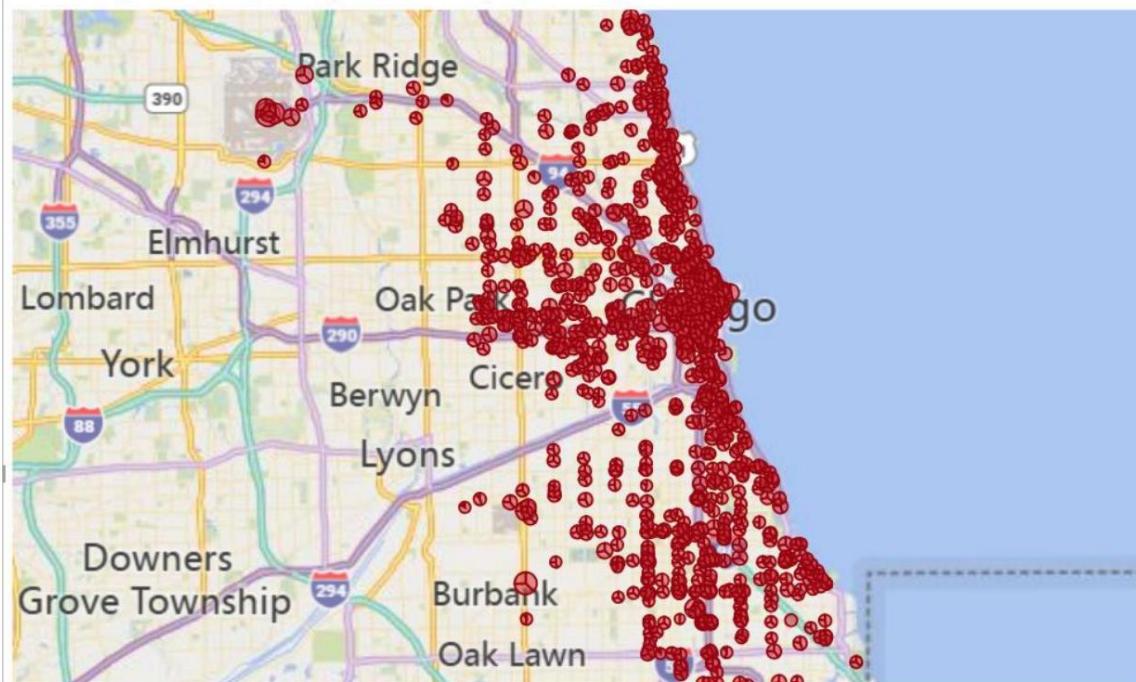
Mapa przestępstw w Chicago w czasie tygodnia (poniedziałek - czwartek)

Weekday Name ● Monday ● Thursday ● Tuesday ● Wednesday



Mapa przestępstw w Chicago w czasie weekendu (piątek - niedziela)

Weekday Name ● Friday ● Saturday ● Sunday



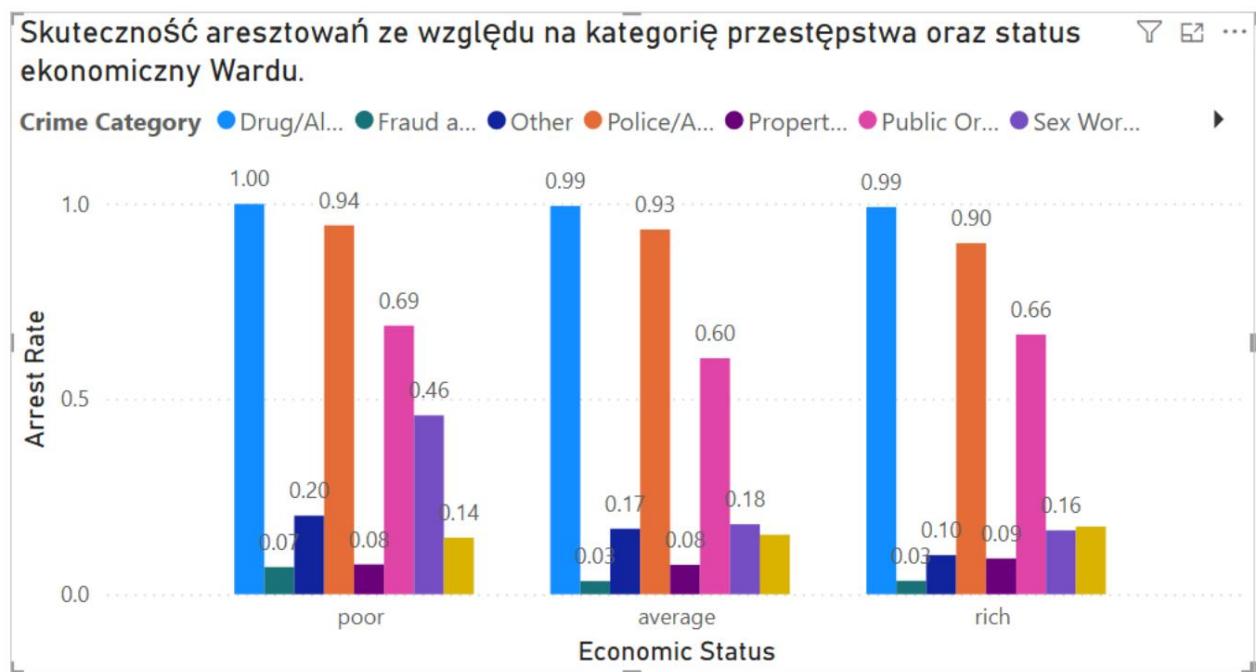
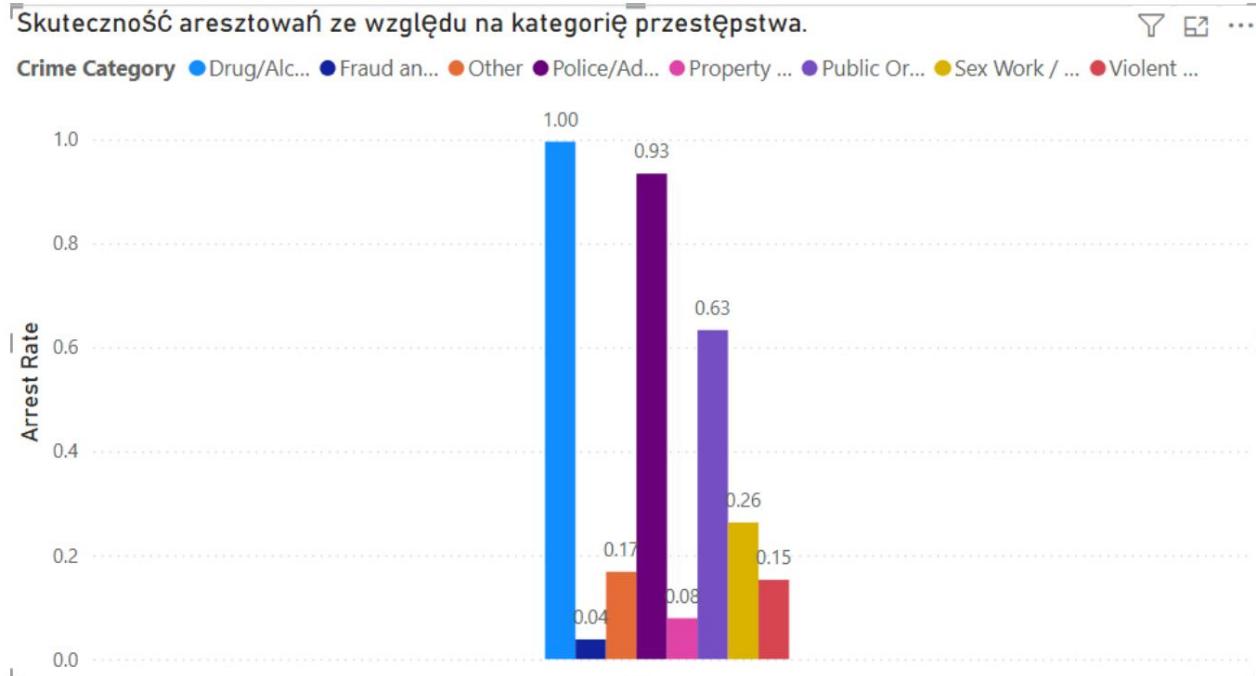
Spatial analysis of crime distribution by day of the week reveals that:

- **The general distribution of crime locations on weekends and weekdays is very similar**
 - crimes are concentrated in the same, most densely populated areas of the city.
- **On weekends, there is a slightly higher incidence of incidents in the city centre and along the waterfront areas,** which may be related to night-time activity, social life and tourism.
- In addition, **most high-crime locations occur both during the week and weekends,** suggesting a constant nature of threats in these areas.

Analysis 8: Number of crimes by type and outcome (arrest)

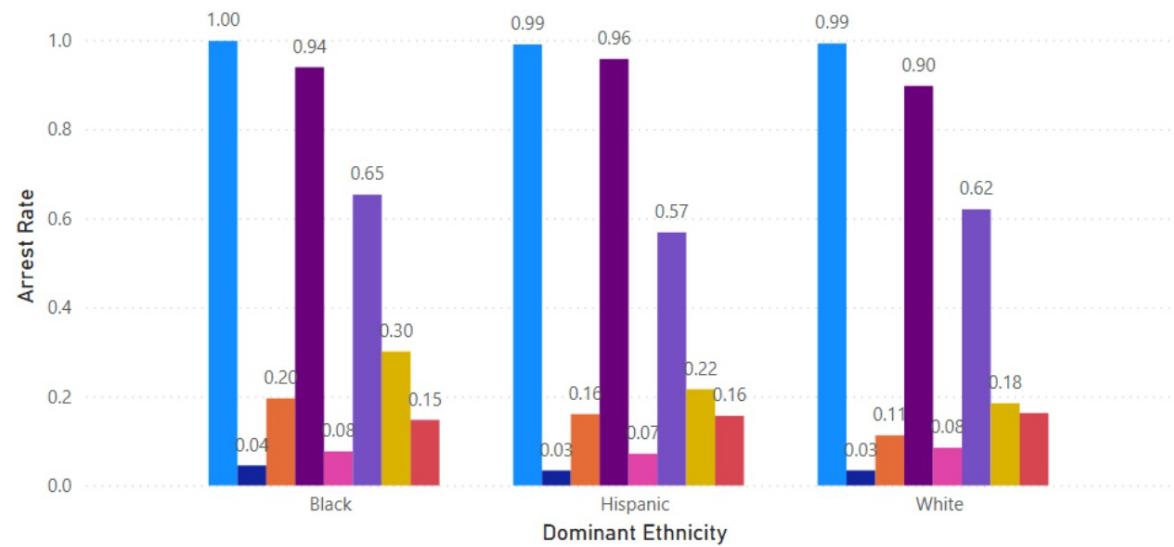
Some analysis has already been done in the analysis of crime outcomes by ward economic status in Study 4, so we will supplement it with the remaining demographic factors.

And we'll delve into Violent Crimes.



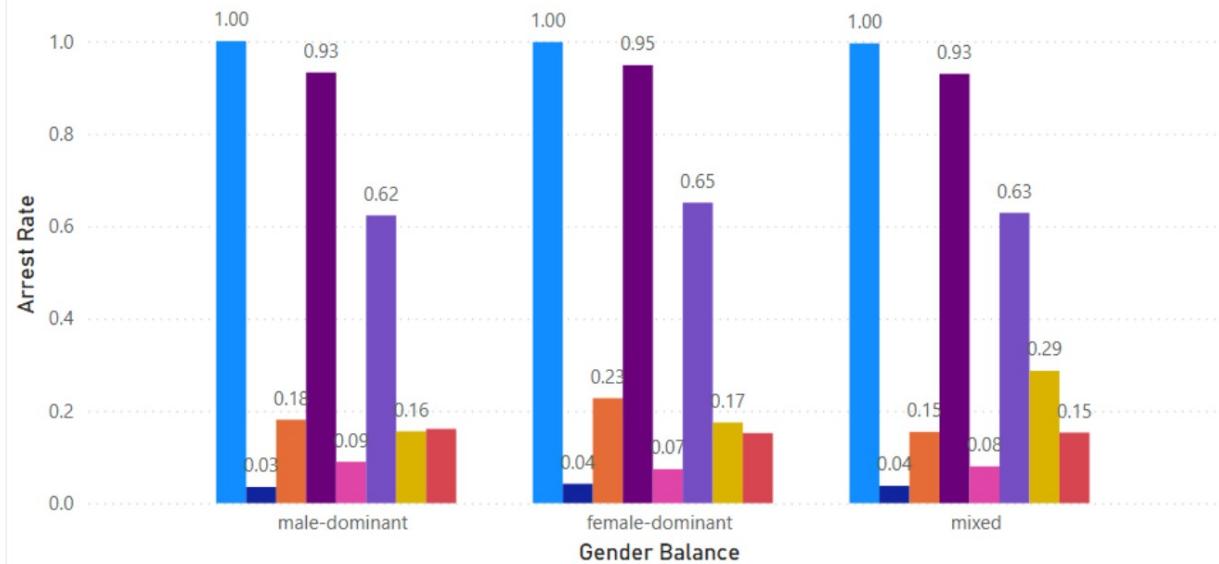
Skuteczność aresztowań ze względu na kategorię przestępstwa i dominującą grupę etniczną warty.

Crime Category ● Drug/Alc... ● Fraud an... ● Other ● Police/Ad... ● Property ... ● Public Or... ● Sex Work / ... ● Violent ...



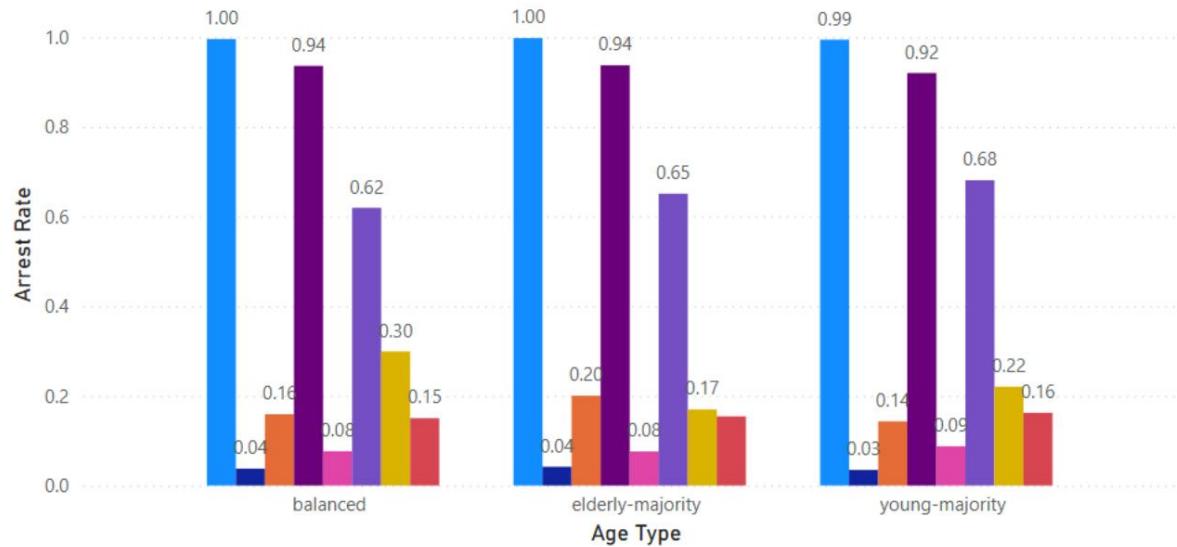
Skuteczność aresztowań ze względu na kategorię przestępstwa i balans płci warty.

Crime Category ● Drug/Alc... ● Fraud an... ● Other ● Police/Ad... ● Property ... ● Public Or... ● Sex Work / ... ● Violent ...



Skuteczność aresztowań ze względu na kategorię przestępstwa i strukturę wiekową.

Crime Category ● Drug/Alc... ● Fraud an... ● Other ● Police/Ad... ● Property ... ● Public Or... ● Sex Work / ... ● Violent ...

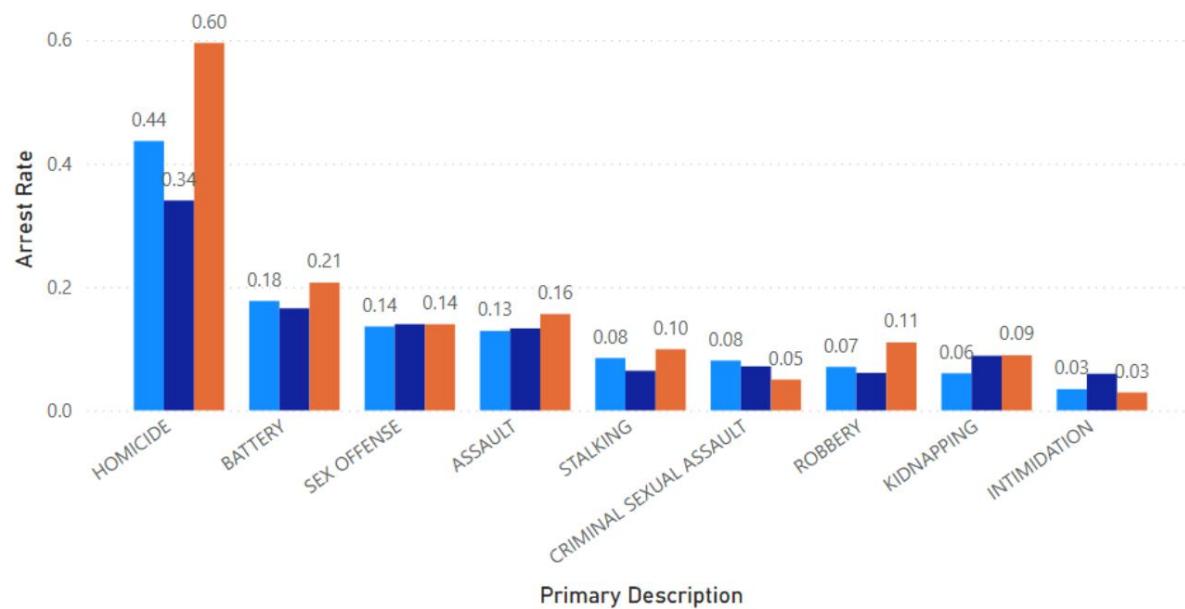


- **Analysis of the effectiveness of arrests by various demographic and economic factors (property status, dominant ethnic group, age structure) did not reveal any significant differences in the way law enforcement agencies operated between wards.**
- **Arrest rates for specific crime categories remain similar across all classifications,** suggesting relatively consistent law enforcement across the city.
- **Exceptions are possible for specific crimes or areas,** however, Due to too high aggregation of data at the level of general categories, **we are unable to unambiguously detect them** in this analysis.
- Therefore, **we will conduct an in-depth analysis of one of the selected categories of offences** to check whether, in the case of a specific type of offence, there are differences in the effectiveness of arrests between wards with different socio-economic profiles.

Arrest rates by violent crime type and ward economic status.

Skuteczność aresztowań według typu przestępstwa z wykorzystaniem przemocy i statusu ekonomicznego wardu.

Economic Status ● average ● poor ● rich



- The highest arrest rate is for homicides (**HOMICIDE**), which is understandable given the priority of their prosecution:

- o In the rich wards it is as much as **59.5%**,
 - in the middle-income group **43.6%**,
 - and among the poor only **33.9%**.

- o This may indicate greater investigative resources or greater law enforcement effectiveness in more affluent neighborhoods.
- For other violent crimes, such as **BATTERY** and **ASSAULT**, the arrest rate is much lower and gradually decreases with lower economic status:

- o **BATTERIES**: from **20.7%** in rich wards to **16.5%** in poor wards.
 - o **ASSAULT**: from **15.6%** (rich) to **13.2%** (poor).

- In the case of sexual offences:

- o The arrest rate for *SEX OFFENSE* is relatively low and **independent of ward wealth** (13–14% in all groups).
- o For *CRIMINAL SEXUAL ASSAULT*, however, there is a significant difference - **8% in middle wards and 7.1% in poor wards**, but only **5% in rich wards**, which may indicate differences in the method of classification or approach of investigators.
- The lowest arrest rates occur for crimes such as:
 - o *INTIMIDATION* (2.8% in rich wards, 5.9% in poor wards),
 - o *KIDNAPPING* and *STALKING* also remain at levels below 10%, regardless of ward status.

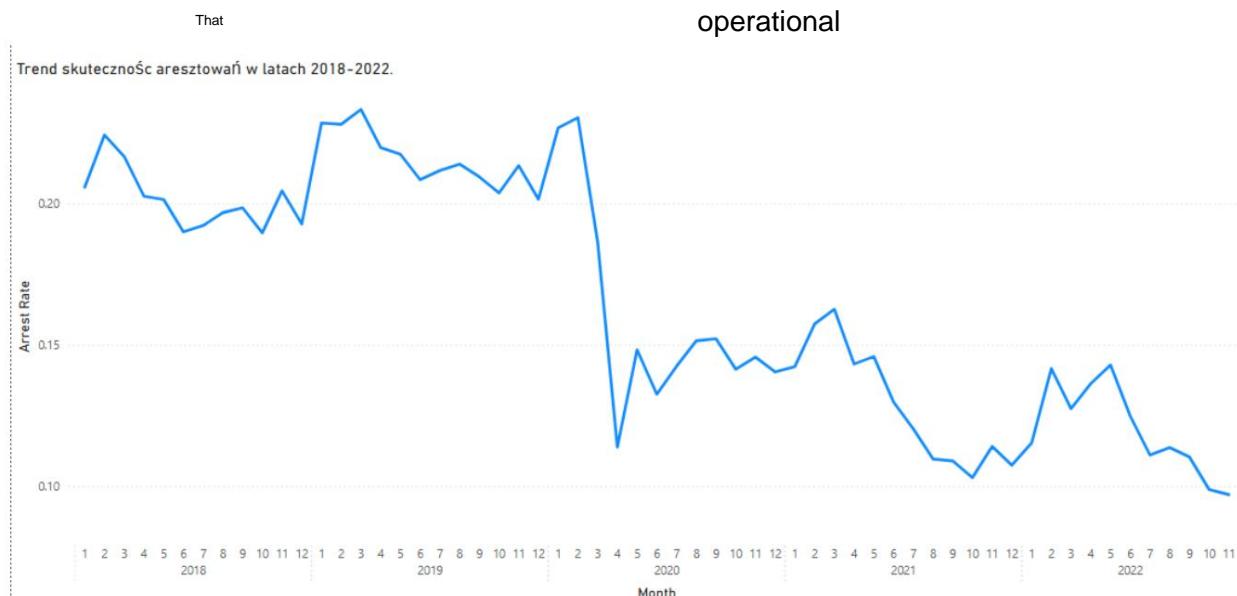
Summary:

In the case of violent crimes:

- **The economic status of a district affects the effectiveness of arrests**, especially for more serious crimes such as *HOMICIDE*.
- In many cases, **the poorer the ward, the lower the apprehension rate**, which may be due to fewer police resources, weaker community cooperation, or a greater burden on law enforcement.
- Differences in arrests may also suggest **systemic inequalities** and it would be worth investigating them further by analysing, for example, investigation times, the rate of unsolved cases or the presence of city surveillance.

Analysis 9: Arrest effectiveness over the years.

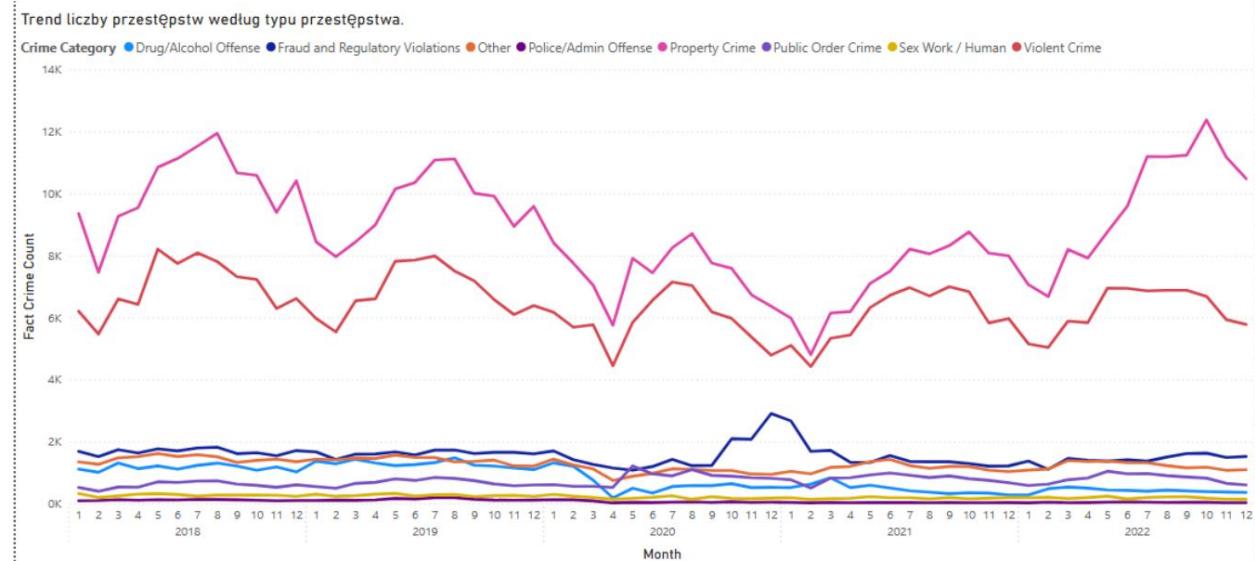
Analysis 9 tracked changes in the Arrest Rate over the years, with particular emphasis on the impact of the COVID-19 pandemic and changes in crime structure. The aim was to determine whether the observed declines in effectiveness were due to systemic or structural changes.



We see a sharp decline in arrest rates in April 2020, with the onset of the COVID-19 lockdown.

Hypotheses to explore:

- The sharp decline in 2020 may be related to:
 - about **pandemic restrictions** affecting the operational capacity of services orderlies,
 - about **changes in the types of crimes** – e.g. an increase in minor offences, where no arrest is made,
 - o **reduced number of direct interventions** or restrictions on detention.

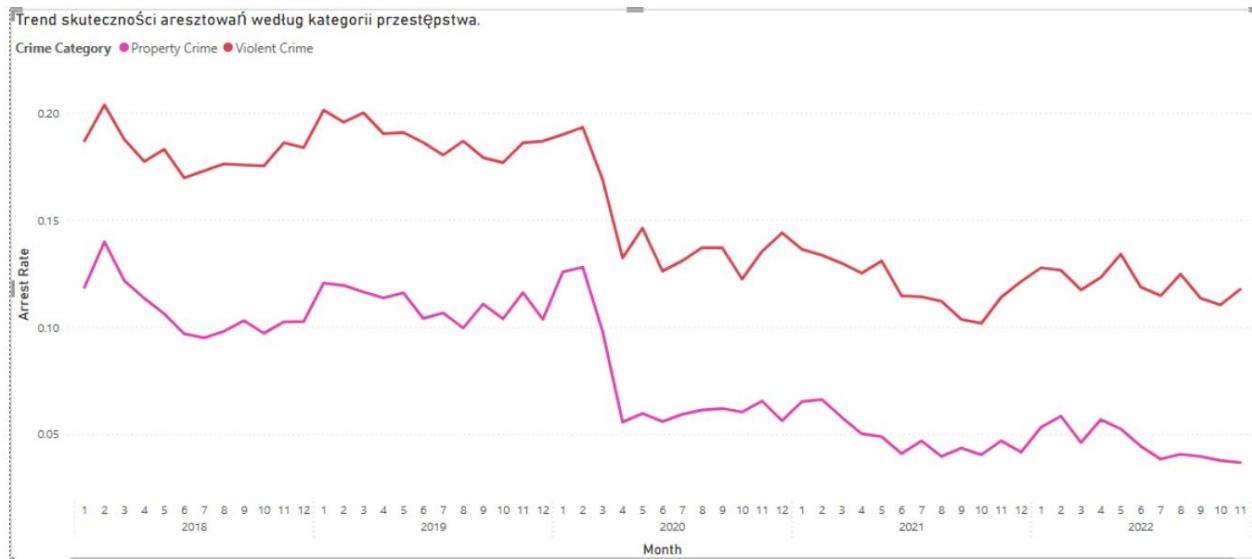
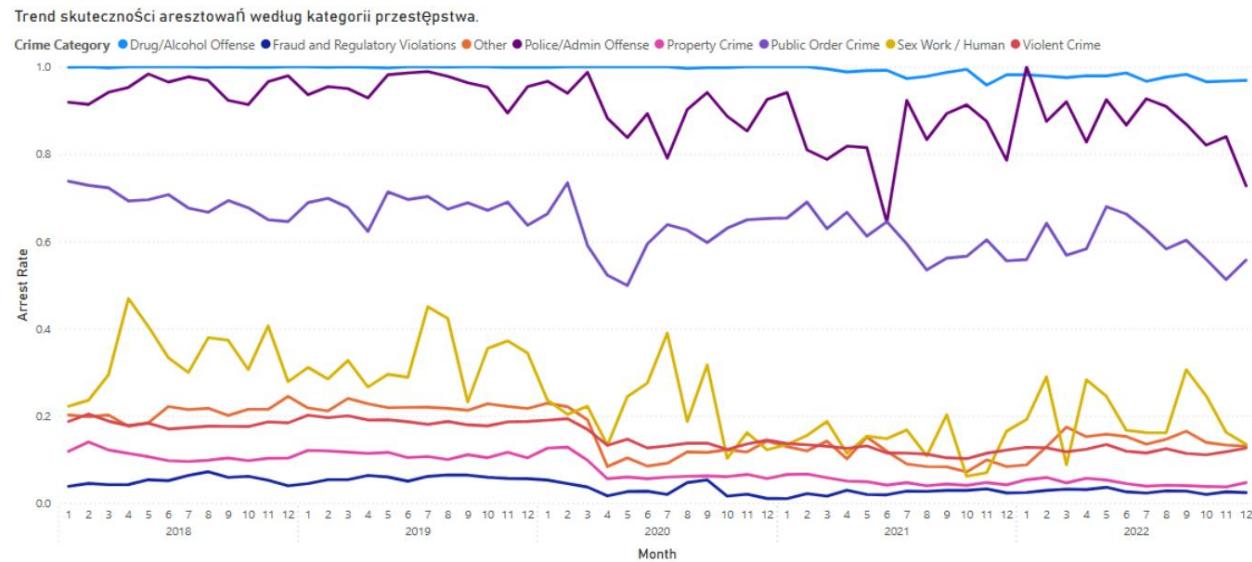


Conclusions:

- In April and May 2020, there was indeed a **significant decline in the number of crimes** in most categories (e.g. *Violent Crime*, *Drug/Alcohol Offense*, *Sex Work*), which coincides with the initial phase of the COVID-19 pandemic and the restrictions introduced.
- After these months, most categories gradually return to pre-pandemic levels, and the number of crimes **in the second half of 2020 and 2021** remains relatively stable.
- **However, there is no significant change in the structure of crime types** – the proportions between categories (e.g. *Violent*, *Property*, *Fraud*) remain fairly constant over time.

years.

So let's analyze effectiveness over the years by crime type.



1. Violent Crime

In 2018-2019, the effectiveness of arrests in this category remained stable, typically in the range of 17-20%. Starting in March 2020, with the onset of the COVID-19 pandemic, a sharp decline in effectiveness was observed:

- April 2020: 13.2%
- June 2020: 12.6%
- July 2021: 11.4%
- October 2021: lowest level recorded – 10.1%

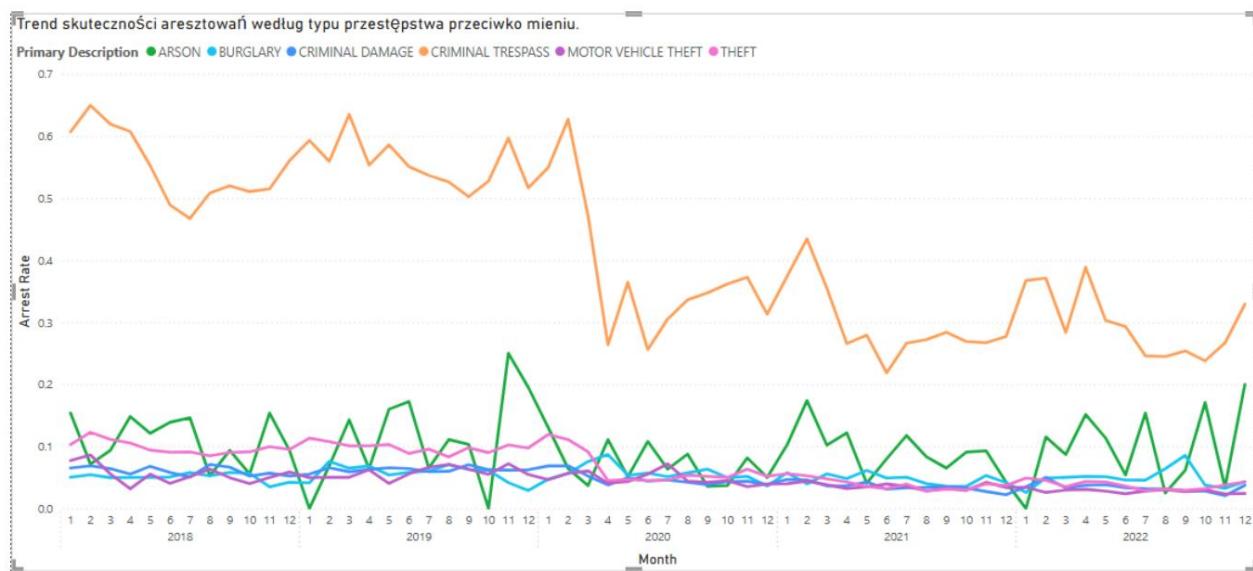
In the following months, this indicator gradually increased, but by the end of 2022 it had not returned to pre-pandemic levels.

2. Property Crime

The arrest rate for this crime category in 2018-2019 hovered around 10-12%. Since March 2020, there has been a sharp decline:

- April 2020: 5.6%
- June 2020: 5.5%
- In 2021–2022: stabilization at a low level, in the range of 3.7–6%

This decline was not only sharp, but also long-lasting.



In 2018-2019, the effectiveness of arrests for the crime of *Criminal Trespass* remained at a relatively high level – most often exceeding 50%. However, since the beginning of 2020, we have observed a clear and lasting decline in this indicator. In the following months, it rarely exceeded 35%, and periodically even dropped below 25%.

Porównanie liczby kradzieży z liczbą wtargnięć na przestrzeni lat 2018-2022.

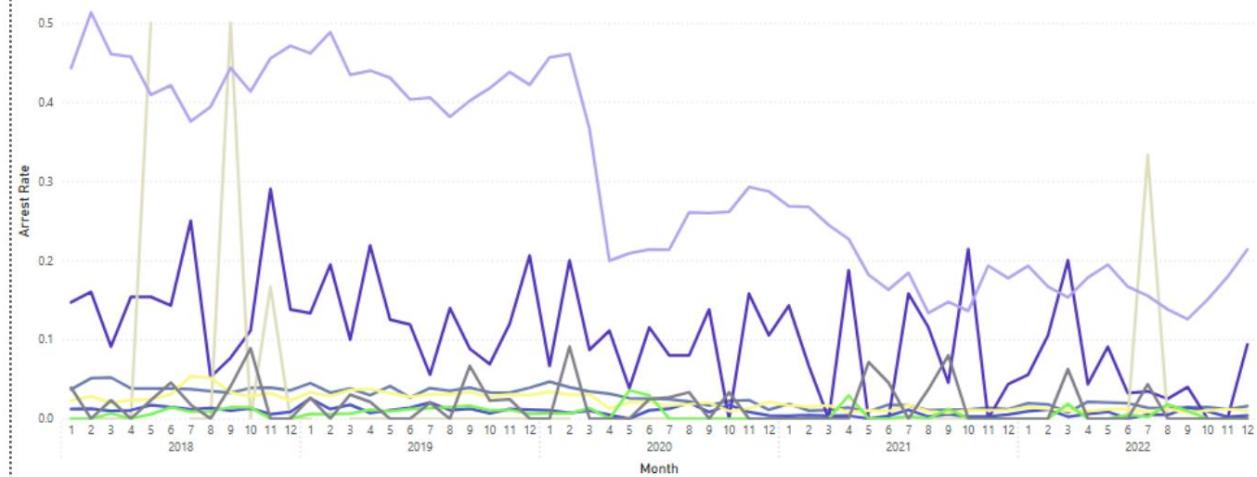
Primary Description ● CRIMINAL TRESPASS ● THEFT



Despite this change, *Criminal Trespass* only constitutes a small percentage of all property crimes. The dominant category in this group is *Theft*, which occurred in the number of several thousand cases each month. It is in this category, especially in the *Retail Theft subtype*, that the most noticeable decrease in the effectiveness of arrests was noted. In 2019, this rate regularly exceeded 40%, but since 2020 it has been gradually decreasing - in some months even falling below 20%.

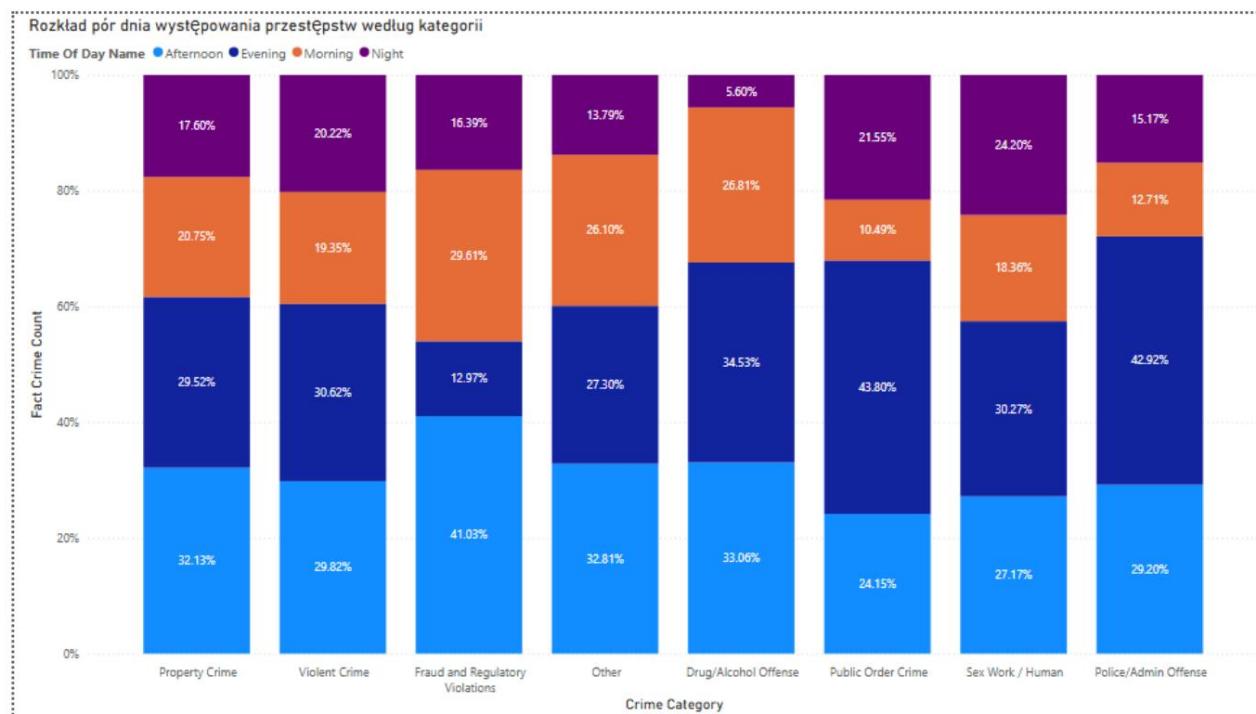
Trend skuteczności aresztowań w przypadku i kradzieży.

Secondary Description ● \$500 AND UNDER ● ATTEMPT THEFT ● DELIVERY CONTAINER THEFT ● FROM BUILDING ● OVER \$500 ● POCKET-PICKING ● PURSE-SNATCHING ● RETAIL THEFT



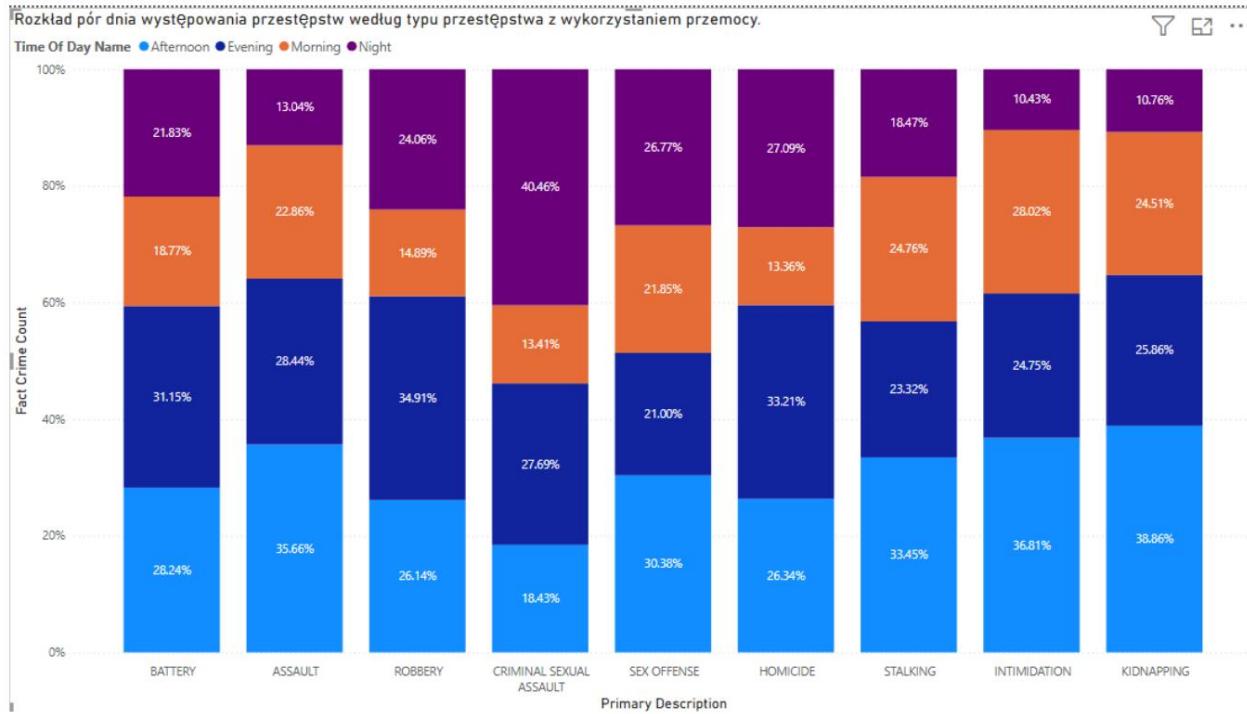
In light of the above, it can be concluded that despite the decline in *Criminal Trespass* effectiveness, it was *Retail Theft* that had a decisive influence on the general trend of lower arrests for property crimes. The reasons for this decline can be seen in operational restrictions of the police during the pandemic, changes in priorities of the services' activities, as well as a possible decrease in the emphasis on prosecuting petty shoplifting during this period.

Analysis 10: Crime category by time of day.



Based on data on the number of crimes broken down by time of day and crime category, clear differences can be observed in the time distribution for individual types of acts:

- **Property Crime** is most often committed **in the afternoon (33.5%)** and **in the evening (30.8%)**, these two times of day together constitute **over 64%** of all such crimes. 21.6% occur in the morning and 18.4% at night.
- **Violent Crime** is more evenly distributed, but also dominates **in the afternoon (30.8%)** and **evening (31.7%)**, which together constitute **over 62%** of cases. Morning and night – approx. 19% each
- **Drug and alcohol-related crimes** are most concentrated **in the evening (33.7%)** and **in the afternoon (32.3%)**, while **at night (5.5%)** they are relatively rare.
- **Prostitution and human trafficking** are most prevalent **in the evening (30.6%)**, but also **in nights (24.5%)** – which distinguishes them from other categories.
- **Order and administrative crimes** also dominate **in the evening (37%)**, but they occur relatively rarely in the morning (11%).



Data analysis reveals interesting variations in crime types depending on time of day.

Some crimes show a strong concentration at specific times of the day:

- **Criminal Sexual Assault** (rape) is most often committed **at night** – as many as **3152 cases**, which is **approx. 39%** of all reported crimes of this type. In comparison: in the evening – 2157 (27%), in the afternoon – 1436 (18%), in the morning – 1045 (13%). This is apparently **the most nocturnal category of violent crime**.
- **Battery** (assault) is dominant **in the evening (69,241)** and **afternoon (62,783)**, which together account for **over 60%** of cases. Fewer incidents were recorded at night (48,536) and in the morning (41,731), suggesting a higher intensity of this type of incident during hours of social activity.
- **Robbery** also peaks **in the evening (14,806)** and **at night (10,206)** – in total, this is **over 54%** of cases. This crime clearly increases in the late hours.
- **Homicide** is less evenly distributed – most cases are reported **in the evening (1,146)** and **at night (935)**, which together constitutes about **60%** all events.
- **Stalking** (harassment) and **Intimidation** (threats) occur more frequently **during the day**, especially **in the afternoon** – 489 and 293 cases, respectively.

General conclusions

- Sexual crimes and the most violent crimes (such as robberies and murders) tend to occur **in the evening and at night**, when social activity takes place with less supervision and limited access to public assistance.
- Crimes such as stalking, intimidation and kidnapping tend to be **more diurnal in nature**, which may indicate they are linked to personal or professional relationships.
- The data confirm the intuition that **night is the time of greatest risk for sexual crimes**, which may be important for preventive measures and planning of police patrols.

5. KPI + Reports

As part of the project, a KPI (Key Performance Indicator) was designed to monitor the share of domestic crimes in the total number of crimes – the so-called **Domestic Crime Rate**. This KPI was calculated as the quotient of the number of crimes marked with the *domestic* flag and the total number of crimes in a given group.

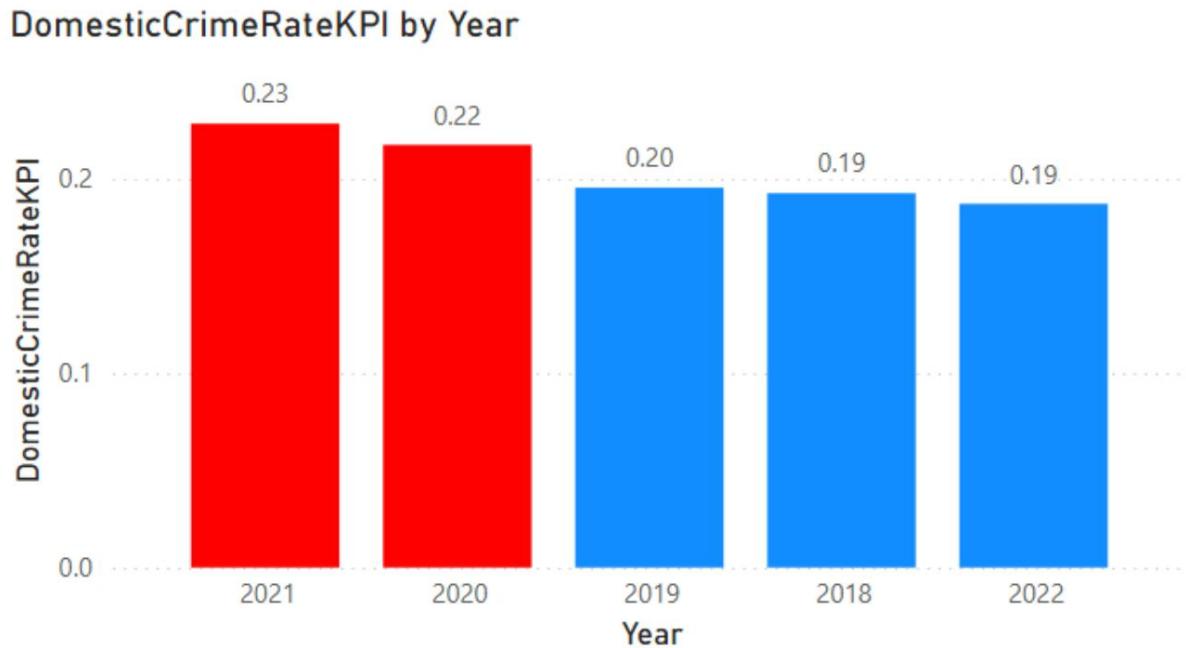
For the purposes of interpreting KPIs, the following scale of threshold values was adopted:

- **KPI value ≤ 0.10** – *Status = 1* (green): The level of domestic crime is considered too low
- **$0.10 < \text{KPI} \leq 0.20$** – *Status = 0* (yellow): Moderate level
- **$\text{KPI} > 0.20$** – *Status = -1* (red): High level, potentially worrisome

Sample breakdown: Domestic Crime Rate by Age Type and Dominant Ethnicity

Age Type	Dominant Ethnicity	DomesticCrimeRateKPI	DomesticCrimeRateKPI Goal	DomesticCrimeRateKPI Status
balanced	Black	0.25	0.10	<input type="radio"/>
balanced	Hispanic	0.21	0.10	<input type="radio"/>
balanced	White	0.13	0.10	<input checked="" type="radio"/>
elderly-majority	Black	0.27	0.10	<input type="radio"/>
elderly-majority	White	0.15	0.10	<input checked="" type="radio"/>
young-majority	Hispanic	0.19	0.10	<input checked="" type="radio"/>
young-majority	White	0.08	0.10	<input checked="" type="radio"/>
Total		0.20	0.10	<input type="radio"/>

Sample breakdown: Domestic Crime Rate by year:



6. Conclusions from Data Analysis:

Based on an OLAP cube built in SSAS, a series of multidimensional analyses were conducted, which allowed for obtaining in-depth insight into the structure and dynamics of crime in Chicago in the years 2018-2022. Integration of factual data with demographic, geographic and temporal dimensions enabled identification of relationships that would be difficult to notice in classic, one-dimensional reports.

Key observations:

- **The crime profile varies by ward economic status** – in poorer wards, violent and drug crimes predominate, while in wealthier wards, financial fraud and offences recorded as “suspicious substances” predominate. This suggests different crime patterns and potential differences in preventive or policing efforts.
- **The dominant ethnic and age group influences the structure of crime** – wards with a predominant young population are characterised by a high share of theft and digital crimes, while older and black wards are more likely to commit domestic crimes.

- **Time and place matter** – many types of crime show clear seasonality (e.g. higher rates in summer) and concentration at certain times of day.
For example, sexual crimes predominate at night, while thefts occur in the afternoon.
- **The COVID-19 pandemic has had a clear impact on the effectiveness of police operations**
 - in 2020 there was a significant decline in the effectiveness of arrests, especially in the categories of property and violent crimes. In the following years, this trend did not return to pre-pandemic levels.
- **The effectiveness of arrests depends on the type of crime, but also on the district**
 - the highest effectiveness is demonstrated by categories of crimes detected directly (e.g. drugs), while for complex crimes (e.g. fraud, sexual violence) the arrest rate is much lower
 - especially in poorer wards.
- **Differences in classifications and investigative approaches** can result in significant discrepancies in statistics – for example, in the Human Trafficking category, cases were recorded only in wealthier wards, which may indicate different approaches or insufficient disclosure of this type of crime in other areas.