

Exercise 1. Business understanding

Project: Crime pattern and trends identifying, its analysis and visualization in the City of Chicago from 2001 and in the City of Los Angeles from 2010 to present.

Team members: Dmytro Tkachuk, Raman Shapaval, Sofiya Demchuk.

Identifying your business goals

- **Background**

The crime rate is changing considerably for the last 20 years. It is important to understand the crime pattern and trend. Data mining helps to recognize dependencies and find the prediction model for location and the probability for the crime.

- **Business goals**

- Based on dataset we will identify and analyze the patterns and trends in crime situation in the City of Chicago (2001 - 2017) and the City of Los Angeles (2010 - 2017).
- Using visualization library and filtering functions we will determine the most “dangerous” districts or the most safe one depending on the day time, period of the year, weekends and etc. We think in this way we can extract benefit for security management in the city. This should help with preventing crimes and consequently with lowering crimes’ rate.
- The dates of crimes can be studied to see crime trends depending on seasons, celebration days and time of the day.
- Based on the data about crimes, we will try to predict exact time of committed crimes depending on type of crime, time and district.

- **Business success criteria**

As our project is not business-oriented, we are planning to do social research. As a result, we aim building good prediction model and in visualization of inspected results described below.

Assessing your situation

- **Inventory of resources:**

- Three students.
- R framework with tutorials (if needed).
- R Shiny framework
- Co-working space.

- **Requirements, assumptions, and constraints:**

- Trello dashboard for the project time-management was created
- CSV data set has been already prepared and lightly tested before actual use
- Final poster with visualized data from which appropriate conclusions could be easily extracted
- **Risks and contingencies:**
 - Time management: contingency plan - define tasks and divide between team members with fixed deadlines/ work hard in the nights.
- **Terminology**

Crime types:

1.1 Criminal trespass - An unlawful intrusion that interferes with one's person or property.

1.2 Battery - the crime or tort of unconsented physical contact with another person, even where the contact is not violent but merely menacing or offensive.

1.3 Arson - the criminal act of deliberately setting fire to property.

1.4 Homicide - the deliberate and unlawful killing of one person by another; murder.

1.5 Obscenity - the state or quality of being obscene; obscene behavior, language, or images.

- **Costs and benefits:**
 - 5 euro for Icecream
 - 2 euro for tea/coffee
 - Approximately 5 euro for snacks

Benefits (above business and social goals) - as all knowledge, are priceless.

Defining your data-mining goals

- **Data-mining goals**
 - Gain skills in working with massive datasets.
 - Improving skills in data preprocessing and filtering, retrieving trends from data, building different kinds of charts depending on type of data.
 - Learn how to visualize data on the maps.
- **Data-mining success criteria:**
 - Extract useful knowledge out of large data sets using algorithms learned from Data Mining course for better manipulation father.

- Calculate the series of measurements to compare models to choose the best for pattern recognition and prediction.
- R Shiny for amazing visualization (maybe some others packages, will be known later).

Exercise 2. Data understanding

Gathering data

- Outline data requirements D
- Verify data availability - data is in free access and already downloaded in csv format for next processing in R or maybe Python.
- Define selection criteria:
 - The data was taken from the U.S. Government's open data: <https://catalog.data.gov/dataset?tags=crime>.
 - Data includes the details about crime in the City of Chicago from 2011 to present and in the city of Los Angeles dating back to 2010. We will use the date, Block, Primary.Type, Description, Location, Arrest, Domestic, District, Ward and Coordinates columns from the dataset for our project.

Describing data

The dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified.

Exploring data

Descriptions of the columns' names in data set of the city of Chicago (the same columns' names used in Los Angeles city

data set):

1. **"ID"** - Unique identifier for the record.
2. **"Case Number"** - The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
3. **"Date"** - Date when the incident occurred. this is sometimes a best estimate.

4. **"Block"** - The partially redacted address where the incident occurred, placing it on the same block as the actual address.
5. **"IUCR"** - The Illinois Uniform Crime Reporting code.
6. **"Primary Type"** - The primary description of the IUCR code.
7. **"Description"** - The secondary description of the IUCR code, a subcategory of the primary description.
8. **"Location Description"** - Description of the location where the incident occurred.
9. **"Arrest"** - Indicates whether an arrest was made.
10. **"Domestic"** - Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
11. **"Beat"** - Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district.
12. **"District"** - Indicates the police district where the incident occurred.
13. **"Ward"** - The ward (City Council district) where the incident occurred.
14. **"FBI Code"** - Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).
15. **"X Coordinate"** and **"Y Coordinate"** - coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
16. **"Year"** - Year the incident occurred.
17. **"Latitude"** and **"Longitude"** - The latitude and longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
18. **"Location"** - The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

Based on above mentioned data columns' names the visualization will be built in the way to allow user discover changes dynamically, changing time of the day, season or year based on own interests.

Verifying data quality:

- ✓ Data exists
- ✓ We can have it without any restrictions, as the resource is open.

✓ Currently dataset seems to be correct as it is extracted from the Chicago Police Department's CLEAR system. Therefore, the Chicago Police Department [does not guarantee](#) (either expressed or implied) the accuracy, completeness, timeliness, or correct sequencing of the information and the information should not be used for comparison purposes over time.

Exercise 3. Setting up and planning your project

We are planning to do three main tasks, where one of them is poster which will be created together using all visualizations generated using code and content. Second task is generating map of crimes using Shiny framework. The third task is to analyze given data, generate visualization and build the model, which will predict exact time of crime, based on given data.

Approximate time spent by each student is 20+ hours/week.