

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	2
1 СОЗДАНИЕ КОНВЕЙЕРА ДЛЯ ПЕРЕДАЧИ ДАННЫХ .....	3
1.1 Структура и описание данных .....	3
1.2 Стек используемых технологий.....	5
2 АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ .....	12
ЗАКЛЮЧЕНИЕ .....	24
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ .....	25
ПРИЛОЖЕНИЯ.....	27

# ВВЕДЕНИЕ

Современное общество сталкивается с рядом социальных проблем, самой серьезной из которых является преступность. Статистика преступлений пугает своими цифрами, и для борьбы с этим негативным явлением необходимо проводить анализ данных, который позволяет выявить основные тенденции, группы риска, причины и мотивы, а также помогает разрабатывать эффективные меры по предотвращению преступлений.

В современном мире объем данных растет с каждым днем, и для эффективного принятия решений важно уметь анализировать и интерпретировать эту информацию. Анализ данных является неотъемлемой частью многих областей жизни, включая бизнес, науку, медицину, образование и правоохранительную деятельность.

Курсовая работа будет сосредоточена на практическом применении статистических методов и демонстрации их использования с помощью программного инструментария. Будет использован язык программирования Python и соответствующие библиотеки для выполнения анализа данных и построения моделей.

Цель данной курсовой работы — проанализировать имеющуюся выборку и проанализировать основные характеристики жертв преступлений, основные виды преступлений, сравнить уровень преступности в разных районах Лос-Анджелеса, оценить эффективность работы правоохранительных органов и оценить динамику преступности во времени.

Из поставленной цели вытекают следующие задачи:

- составить конвейер для сбора и передачи данных;
- определить ключевые вопросы для проведения анализа;
- провести анализ в соответствии с целью;
- визуализировать полученные результаты.

# 1 СОЗДАНИЕ КОНВЕЙЕРА ДЛЯ ПЕРЕДАЧИ ДАННЫХ

## 1.1 Структура и описание данных

Набор данных представляет собой полный отчет о преступных инцидентах в городе Лос-Анджелес с 2020 года. Данные взяты из оригинальных отчетов о преступлениях, которые изначально были написаны на бумаге, что может привести к некоторым неточностям. Файл представлен в формате csv. В наборе данных содержится 28 признаков и более 900000 записей с данными о преступлениях.

Данные содержат следующие атрибуты:

- DR\_NO — номер записи: официальный файл состоит из 2-значного года, ID области и 5 цифр;
- DATE RPTD — дата подачи заявления: MM/DD/YYYY;
- DATE OCC — дата происшествия: MM/DD/YYYY;
- TIME OCC — время происшествия в 24-часовом военном формате;
- AREA — район: LAPD имеет 21 район полиции, известных как географические районы в отделеб, эти географические районы пронумерованы от 1 до 21;
- AREA NAME — название района: 21 географический район или патрульное подразделение также получают обозначение по имени, которое ссылается на ориентир или окружающее сообщество, за которое оно отвечает;
- Rpt Dist No — четырёхзначный код, отображающий под-район в географическом районе;
- Crm Cd — показывает совершенное преступление;
- Crm Cd Desc — определяет предоставленный код преступления;

- Mocodes — модус операнди: деятельность, связанная с подозреваемым при совершении преступления;
- Vict Age — возраст жертвы (двухзначное числовое значение);
- Vict Sex — пол жертвы (F — Женщина, M — Мужчина, X — Неизвестно);
- Vict Descent — код происхождения: A — Другие азиаты, B — Чёрные, C — Китайцы, D — Камбоджийцы, F — Филиппинцы, G — Гуамцы, H — Испанцы/латиноамериканцы/мексиканцы, I — Американские индейцы/инуиты, J — Японцы, K — Корейцы, L — Лаосцы, O — Другие, P — Тихоокеанские острова, S — Самоанцы, U — Гавайцы, V — Вьетнамцы, W — Белые, X — Неизвестно, Z — Индусы;
- Premis Cd — тип структуры, транспортного средства или места, где произошло преступление;
- Premis Desc — определяет предоставленный код места;
- Weapon Used Cd — тип оружия, использованного при преступлении;
- Weapon Desc — определяет предоставленный код использованного оружия;
- Status — статус дела (IC — значение по умолчанию);
- Status Desc — определяет предоставленный код статуса;
- Crm Cd 1 — указывает совершенное преступление (код преступления 1 является основным и наиболее серьезным, коды преступлений 2, 3 и 4 соответственно являются менее серьёзными нарушениями);
- Crm Cd 2 — может содержать код для дополнительного преступления, менее серьёзного, чем код преступления 1;
- Crm Cd 3 — может содержать код для дополнительного преступления, менее серьёзного, чем код преступления 2;
- Crm Cd 4 — может содержать код для дополнительного преступления, менее серьёзного, чем код преступления 3;
- LOCATION — улица преступного инцидента, округлённая до

ближайшего сотенного блока для защиты анонимности;

- Cross Street — перекрёстная улица округленного адреса;
- LAT — широта;
- LON — долгота.

## **1.2 Стек используемых технологий**

В настоящее время в области анализа данных доступно множество различных инструментов. В данном исследовании планируется использовать инструментарий, специализированный на обработке и хранении больших объемов данных.

Для хранения и обработки данных в проекте будет задействованы MariaDB и Apache Hive. MariaDB представляет собой продвинутую систему управления реляционными базами данных, которая заимствует и расширяет функциональность MySQL. Отличительные особенности MariaDB включают в себя повышенную производительность, улучшенные возможности управления данными и более надежный код, что делает ее более привлекательной для разработчиков. Кроме того, MariaDB внедряет более эффективный оптимизатор запросов и безопасные индексы, что обеспечивает более устойчивые алгоритмы хранения информации [1.1].

Apache Hive, в свою очередь, представляет собой SQL-интерфейс для работы с данными в рамках платформы Apache Hadoop. Hive позволяет выполнять запросы, агрегировать и анализировать данные в формате SQL, что делает работу с данными более удобной и интуитивно понятной [1.2]. С помощью HiveQL, запросы переводятся в Java-код задач MapReduce, что обеспечивает эффективную обработку данных в распределенной файловой системе HDFS.

Использование MariaDB и Apache Hive позволит эффективно хранить и анализировать данные, особенно в условиях работы с распределенной файловой

системой HDFS. Для передачи данных между HDFS и MariaDB будет применяться Apache Sqoop, а для моделирования потоков передачи данных - Apache Kafka. Apache Kafka дает возможность собирать, хранить и распределять данные между компонентами приложения, обеспечивая высокую доступность и надежность. Данные из MariaDB будут поступать в Kafka через Flume, что позволит эффективно управлять потоками данных и направлять их на нужные пункты назначения [1.3].

Также будет задействован Apache Spark — фреймворк с открытым исходным кодом для реализации распределённой обработки данных. Он предоставит возможности для распределенного и параллельного выполнения задач на кластерах компьютеров [1.4].

PySpark позволит использовать преимущества Apache Spark, используя синтаксис языка Python и его функции. Он обеспечивает доступ к мощным функциям Spark для распределенной обработки данных.

Все эти технологии совместно образуют мощный и эффективный стек инструментов, который дает возможность эффективно обрабатывать, хранить и анализировать данные в работе, открывая новые возможности для анализа информации и принятия важных решений на основе данных [1.5].

Перед проведением анализа и реализацией конвейера, была разработана его схема, показанная на Рисунке 1.1.

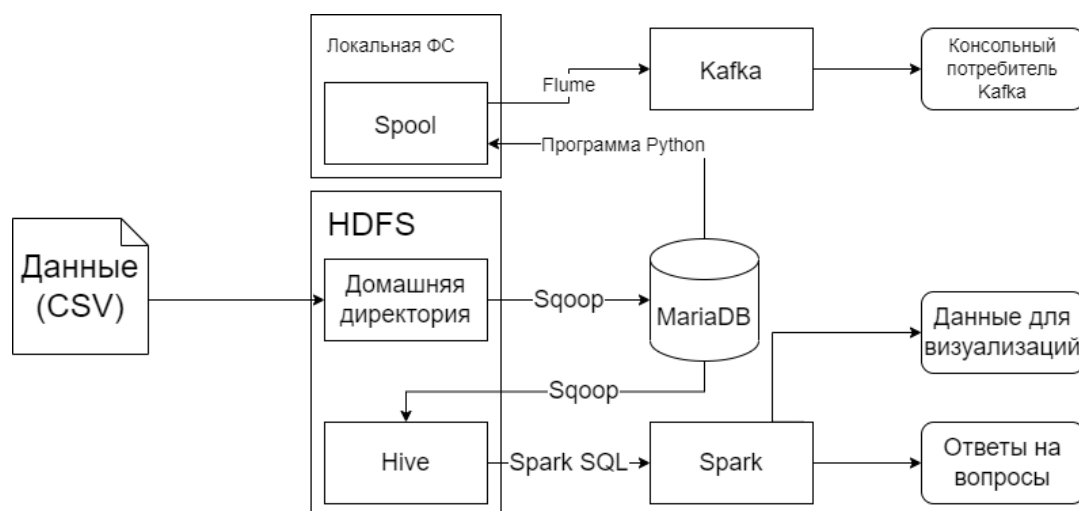


Рисунок 0.1 — Схема конвейера данных

```

CREATE TABLE [dbo].[crimes]
(
    [CrimeID] int NOT NULL,
    [Crime Desc] varchar(100) NOT NULL,
    [Modus] varchar(100) NOT NULL,
    [Vict Age] varchar(100) NOT NULL,
    [Vict Sex] varchar(100) NOT NULL,
    [Vict Desc] varchar(100) NOT NULL,
    [Premis Desc] varchar(100) NOT NULL,
    [Premis Desc] varchar(100) NOT NULL,
    [Weapon Used] varchar(100) NOT NULL,
    [Weapon Desc] varchar(100) NOT NULL,
    [Status Desc] varchar(100) NOT NULL,
    [Crme Cd 1] varchar(100) NOT NULL,
    [Crme Cd 2] varchar(100) NOT NULL,
    [Crme Cd 3] varchar(100) NOT NULL,
    [Crme Cd 4] varchar(100) NOT NULL,
    [Location] varchar(100) NOT NULL,
    [Cross Street] varchar(100) NOT NULL,
    [Latitude] float NOT NULL,
    [Longitude] float NOT NULL
)
GO

```

Далее первоначальный файл с данными был экспортирован в MariaDB через инструмент Apache Sqoop. Команда для экспорта представлена на Рисунке 1.3.

### Рисунок 0.3 — Экспорт данных в базу данных

DR No	Date Rptd	Date Occ	Time Occ	Area	Area Name	Rpt_Dist No	Part	Crm Cd	Crm Cd Desc	Modcodes	Vict Age	Vict Sex	Vict Descnt
Premis Cd	Premis Desc	Weapon_Used_Cd	Weapon_Desc			Status	Status_Desc	Crm_Cd_1	Crm_Cd_2	Crm_Cd_3	Crm_Cd_4	Location	
Street	Lat	Lon											
220312085	06/11/2022 12:00:00 AM	06/10/2022 12:00:00 AM	1930	03	Southwest	0363	1	510	VEHICLE - STOLEN				
108	PARKING LOT					IC	Invest Cont	510			4000	CRENSHAW	BL
	34.012	-118.3351											
228916403	11/21/2022 12:00:00 AM	11/20/2022 12:00:00 AM	1830	09	Van Nuys	0911	1	519	VEHICLE - STOLEN				
101	STREET					IC	Invest Cont	510			6500	LANGDON	AV
	34.189	-118.4692											
22812950	08/05/2022 12:00:00 AM	08/05/2022 12:00:00 AM	1520	08	West LA	0839	1	442	SHOPLIFTING - PETTY THEFT (\$950 & UNDER)	0325 2004		X	X
404	DEPARTMENT STORE					IC	Invest Cont	442			10200	SANTA MONICA	BL
	34.0611	-118.4184											
22151038	06/13/2022 12:00:00 AM	06/13/2022 12:00:00 AM	1320	15	N Hollywood	1504	1	510	VEHICLE - STOLEN				
108	PARKING LOT					AA	Adult Arrest	510			7300	LAKERSHIM	BL
	34.2945	-118.392											
226707950	04/13/2022 12:00:00 AM	04/13/2022 12:00:00 AM	2000	07	Wilshire	0734	1	480	BIKE - STOLEN	0346		57	M
102	SIDEMALK					IC	Invest Cont	480			6300 W 3RD		ST
	34.0713	-118.3595											
221415693	08/24/2022 12:00:00 AM	08/24/2022 12:00:00 AM	1500	14	Pacific	1483	2	354	THEFT OF IDENTITY	0377		76	F
501	SINGLE FAMILY DWELLING					IC	Invest Cont	354			8600	FALMOUTH	AV
	33.9572	-118.4363											
220215346	08/27/2022 12:00:00 AM	08/27/2022 12:00:00 AM	1000	02	Rampart	0202	1	421	THEFT FROM MOTOR VEHICLE - ATTEMPT	1022 1300 1501	51	F	H
108	PARKING LOT	400	STRONG-ARM (HANDS, FIST, FEET OR BODYLY FORCE)			IC	Invest Cont	421				VERMONT	LOC
222109071	05/31/2022 12:00:00 AM	05/30/2022 12:00:00 AM	2000	21	Topanga	2155	1	442	SHOPLIFTING - PETTY THEFT (\$950 & UNDER)	0325		35	M
402	MARKET					IC	Invest Cont	442			6200	TOPANGA CANYON	BL
	34.1829	-118.6059											
221110075	06/26/2022 12:00:00 AM	06/25/2022 12:00:00 AM	1830	11	Northeast	1151	1	330	BURGLARY FROM VEHICLE	1022 1300 0344	0	F	F
101	STREET					IC	Invest Cont	330				NORMANDIE	FOU
	34.0960	-118.3005											
221704120	01/05/2022 12:00:00 AM	01/04/2022 12:00:00 AM	1655	17	Devonshire	1791	1	510	VEHICLE - STOLEN				
108	PARKING LOT					IC	Invest Cont	510			20900	BRYANT	ST
	34.227	-118.5885											

10 rows in set (0.00 sec)

Далее необходимо изменить типы данных столбцов. Первоначально

создавалась таблица с одинаковым типом данных для всех столбцов, однако в ней присутствуют не только строки, но и другие типы данных. Поэтому для корректной обработки и анализа необходимо в столбцах с числовой информацией изменить тип данных со сточного на целочисленный (Vict\_age) и с плавающей точкой (Lat, Lon), а также дату/время (Date\_Rptd, Date\_Occ, Time\_Occ). Далее необходимо уменьшить хранимые размеры строк в остальных столбцах и добавить новый столбец, являющийся первичным ключом. Результат произведенных изменений представлен на Рисунке 1.5.

```
MariaDB [crimes]> desc crimes_data;
```

Field	Type	Null	Key	Default	Extra
Id	int(11)	NO	PRI	NULL	auto_increment
Dr_No	varchar(10)	YES		NULL	
Date_Rptd	datetime	YES		NULL	
Date_Occ	datetime	YES		NULL	
Time_Occ	time	YES		NULL	
Area	char(2)	YES		NULL	
Area_Name	varchar(20)	YES		NULL	
Rpt_Dist_No	char(4)	YES		NULL	
Part	char(1)	YES		NULL	
Crn_Cd	char(3)	YES		NULL	
Crn_Cd_Desc	varchar(100)	YES		NULL	
Mocodes	varchar(100)	YES		NULL	
Vict_Age	int(3)	YES		NULL	
Vict_Sex	char(1)	YES		NULL	
Vict_Descent	char(1)	YES		NULL	
Premis_Cd	varchar(5)	YES		NULL	
Premis_Desc	varchar(100)	YES		NULL	
Weapon_Used_Cd	varchar(5)	YES		NULL	
Weapon_Desc	varchar(100)	YES		NULL	
Status	char(2)	YES		NULL	
Status_Desc	varchar(30)	YES		NULL	
Crn_Cd_1	char(3)	YES		NULL	
Crn_Cd_2	char(3)	YES		NULL	
Crn_Cd_3	char(3)	YES		NULL	
Crn_Cd_4	char(3)	YES		NULL	
Location	varchar(100)	YES		NULL	
Cross_Street	varchar(100)	YES		NULL	
Lat	double(10,6)	YES		NULL	
Lon	double(10,6)	YES		NULL	

29 rows in set (0.00 sec)

Рисунок 0.5 — Таблица в базе данных после ее изменения

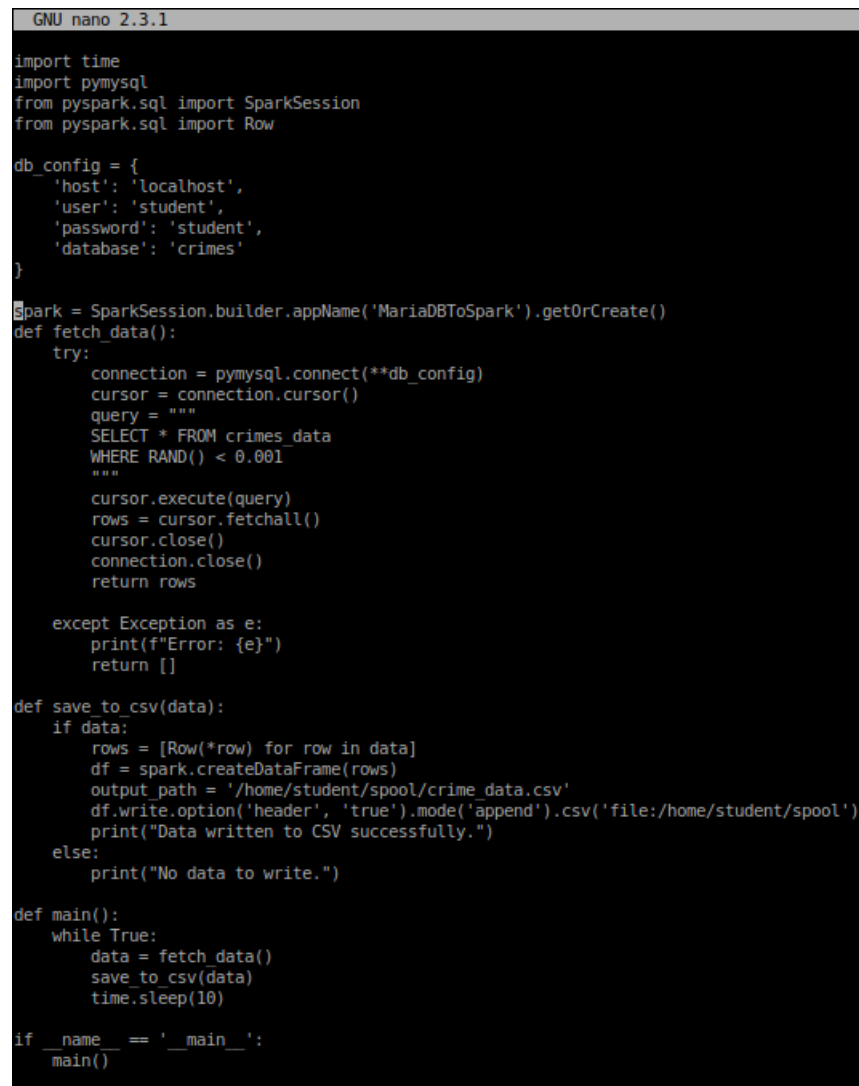
Далее для создания потока данных был создан новый топик с помощью Apache Kafka. Результат создания представлен на Рисунке 1.6.

```
[student@localhost ~]$ kafka-topics --create --bootstrap-server localhost:9092 --replication-factor 1 --partitions 1 --topic crime topic
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not both.
Created topic crime topic.
[student@localhost ~]$ kafka-topics --list --bootstrap-server localhost:9092
consumer offsets
crime topic
```

Рисунок 0.6 — Создание топика Kafka



Для создания потока данных был написан скрипт. Первоначально происходит подключение к MariaDB и проверка соединения. Далее выполняется запрос на выборку 5 % имеющихся данных. Из них, с помощью инструмента spark, формируется датафрейм и записывается в формате csv в папку spool в HDFS. Эти операции происходят каждые 10 секунд. Написанный скрипт представлен на Рисунке 1.7



```
GNU nano 2.3.1

import time
import pymysql
from pyspark.sql import SparkSession
from pyspark.sql import Row

db_config = {
    'host': 'localhost',
    'user': 'student',
    'password': 'student',
    'database': 'crimes'
}

spark = SparkSession.builder.appName('MariaDBToSpark').getOrCreate()
def fetch_data():
    try:
        connection = pymysql.connect(**db_config)
        cursor = connection.cursor()
        query = """
        SELECT * FROM crimes_data
        WHERE RAND() < 0.001
        """
        cursor.execute(query)
        rows = cursor.fetchall()
        cursor.close()
        connection.close()
        return rows

    except Exception as e:
        print(f"Error: {e}")
        return []

def save_to_csv(data):
    if data:
        rows = [Row(*row) for row in data]
        df = spark.createDataFrame(rows)
        output_path = '/home/student/spool/crime_data.csv'
        df.write.option('header', 'true').mode('append').csv('file:/home/student/spool')
        print("Data written to CSV successfully.")
    else:
        print("No data to write.")

def main():
    while True:
        data = fetch_data()
        save_to_csv(data)
        time.sleep(10)

if __name__ == '__main__':
    main()
```

**Рисунок 0.7 — Скрипт для создания файлов**

Для передачи данных через агента Flume был написан конфигурационный файл. Его состав представлен на Рисунке 1.8.

```

GNU nano 2.3.1
agent1.sources = src1
agent1.channels = ch1 ch2
agent1.sinks = sink1 sink2

agent1.sources.src1.type = spooldir
agent1.sources.src1.spoolDir = /home/student/spool

agent1.channels.ch1.type = memory
agent1.channels.ch1.capacity = 10000
agent1.channels.ch1.transactionCapacity = 100

agent1.channels.ch2.type = memory
agent1.channels.ch2.capacity = 10000
agent1.channels.ch2.transactionCapacity = 100

agent1.sinks.sink1.type = org.apache.flume.sink.kafka.KafkaSink
agent1.sinks.sink1.kafka.bootstrap.servers = localhost:9092
agent1.sinks.sink1.kafka.topic = crime_topic
agent1.sinks.sink1.kafka.flumeBatchSize = 5
agent1.sinks.sink1.channel = ch1

agent1.sinks.sink2.type = logger
agent1.sinks.sink2.channel = ch2

agent1.sources.src1.channels = ch1 ch2

```

**Рисунок 0.8 — Конфигурационный файл агента Flume**

После запуска агента Flume и написанного скрипта, данные начали поступать в обработчик. Далее можем убедиться, что поступают именно первоначальные данные. Это показано на Рисунке 1.9.

```

2024-05-27 04:45:46,583 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 5F 31 2C 5F 32 2C 5F 33 2C 5F 34 2C 5F 35 2C 5F _1_2_3_4_5_ }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 39 38 38 38 39 2C 32 30 31 32 30 34 34 36 38 2C 98889,201204468, }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 5F 31 2C 5F 32 2C 5F 33 2C 5F 34 2C 5F 35 2C 5F _1_2_3_4_5_ }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 32 34 32 37 37 39 2C 32 31 30 36 30 34 36 30 33 242779,210604603 }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 32 37 38 39 36 38 2C 32 31 30 39 31 32 35 33 38 278968,210912538 }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 5F 31 2C 5F 32 2C 5F 33 2C 5F 34 2C 5F 35 2C 5F _1_2_3_4_5_ }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 33 35 31 39 31 33 2C 32 31 31 36 31 31 36 38 39 351913,211611689 }
2024-05-27 04:45:46,585 [SinkRunner-PollingRunner-DefaultSinkProcessor] INFO org.apache.flume.sink.LoggerSink - Event: { headers:{} body: 34 38 38 38 37 34 2C 32 32 30 38 31 31 30 32 36 488874,220811026 }

```

**Рисунок 0.9 — Логирувание полученных данных в консоль**

После запуска потребителя Kafka, данные начали поступать и в него, что показано на Рисунке 1.10.

```

933951,241809642,2024-05-01 00:00:00,2024-05-01 00:00:00,07:06:00,18,Southeast,1861,2,354,THEFT OF IDENTITY,0100 1822 0922 0930,20,F,H,501,SINGLE FAMILY DWELLING,"",",",IC,Invest Cont,354,"",",",700 W IMPERIAL HY,"",33.931000,-118.287000
935393,241905938,2024-02-20 00:00:00,2024-02-19 00:00:00,08:16:00,19,Mission,1994,2,930,CRIMINAL THREATS - NO WEAPON DISPLAYED,0443 1906 0913 1817 2003,50,F,W,701,HOSPITAL,511,VERBAL THREAT,IC,Invest Cont,930,"",",",14800 ROSCOE BL,"",34.221500,-118.455300
936728,241908003,2024-04-18 00:00:00,2024-04-17 00:00:00,13:30:00,19,Mission,1955,2,946,OTHER MISCELLANEOUS CRIME,"",58,F,H,501,SINGLE FAMILY DWELLING,"",",",IC,Invest Cont,946,"",",",14600 KINGSBURY ST,"",34.263500,-118.450200
937532,242004256,2024-01-07 00:00:00,2024-01-07 00:00:00,10:07:00,20,Olympic,2015,1,310,BURGLARY,1822 1202 0358 0352 0344 0321,65,F,H,502,"MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)",",",",IC,Invest Cont,310,"",",",100 N KENMORE AV,"",34.072800,-118.296500
939072,242006325,2024-02-20 00:00:00,2024-02-20 00:00:00,04:00:00,20,Olympic,2031,1,330,BURGLARY FROM VEHICLE,0344 1307 1609 1822,52,M,F,501,SINGLE FAMILY DWELLING,"",",",IC,Invest Cont,330,"",",",700 S BRONSON AV,"",34.060700,-118.318400
939696,242007182,2024-03-14 00:00:00,2024-03-14 00:00:00,02:10:00,20,Olympic,2004,1,210,ROBBERY,1205 2042 1822 0342 0213 0302 0344 0377 0400 0306,23,M,B,502,"MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)",102,HAND GUN,IC,Invest Cont,210,"",",",500 N HARVARD BL,"",34.079900,-118.306800
940101,242007760,2024-03-26 00:00:00,2024-03-26 00:00:00,14:00:00,20,Olympic,2039,2,900,VIOLATION OF COURT ORDER,2038 2004,16,M,H,502,"MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)",",",",A0,Adult Other,900,"",",",2800 W 8TH ST,"",34.059800,-118.286500
940658,242008735,2024-05-06 00:00:00,2024-05-04 00:00:00,18:14:00,20,Olympic,2029,2,354,THEFT OF IDENTITY,0100 0929 0930 1822,32,M,K,502,"MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)",",",",IC,Invest Cont,354,"",",",400 S VIRGIL AV,"",34.066700,-118.287100
941531,242104998,2024-01-25 00:00:00,2024-01-25 00:00:00,15:30:00,21,Topanga,2105,2,626,INTIMATE PARTNER - SIMPLE ASSAULT,2000 0416,51,M,B,726,POLICE FACILITY,400,"STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)",IC,Invest Cont,626,"",",",21500 SCHOENBORN ST,"",34.220900,-118.598900
942186,242106068,2024-02-15 00:00:00,2023-01-01 00:00:00,18:00:00,21,Topanga,2157,2,626,INTIMATE PARTNER - SIMPLE ASSAULT,2000 0419,-1,F,H,501,SINGLE FAMILY DWELLING,400,"STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)",A0,Adult Other,626,"",",",20700 SKOURAS DR,"",34.192500,-118.584800
942453,242106483,2024-03-01 00:00:00,2024-02-26 00:00:00,08:00:00,21,Topanga,2157,2,626,INTIMATE PARTNER - SIMPLE ASSAULT,2000 0400 1813,37,F,H,502,"MULTI-UNIT DWELLING (APARTMENT, DUPLEX, ETC)",400,"STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)",IC,Invest Cont,626,"",",",21200 KITTRIDGE ST,"",34.191300,-118.594200

```

**Рисунок 0.10 — Запуск потребителя Kafka**

После создания потока данных, импортируем таблицу из MariaDB в сервис Apache Hive для дальнейшего анализа и обработки. Команда для импорта представлена на Рисунке 1.11.

```

[student@localhost ~]$ sqoop import -Dorg.apache.sqoop.splitter.allow_text_splitter=true --connect jdbc:mysql://localhost:3306/crimes --username student --password student --table crimes_data --hive-import --hive-table hive_crimes

```

**Рисунок 0.11 — Импорт данных в Apache Hive**

В результате получаем исходную таблицу, импортированную в HIVE, что показано на Рисунке 1.12.

```

0: jdbc:hive2://> show tables
. . . . . > ;
OK
+-----+
| tab_name |
+-----+
| authors_parquet2 |
| hive_crimes |
| products |
+-----+
3 rows selected (0.774 seconds)

```

**Рисунок 0.12 — Результат импорта**

## 2 АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ

Для начала импортируем в Spark данные из Hive. Код для импорта приведен в Листинге 2.1.

Листинг 0.1 — Импорт данных в Spark из Hive

```
crime_data = spark.read.table("default.hive_crimes")
crime_data.show(5)
```

На Рисунке 2.1 представлен полученный набор данных.

```
+---+-----+-----+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+-----+-----+-----+
| id|   dr_no|       date_rptd|       date_occ|time_occ|area| area_name|rpt_dist|
no|part|crm_cd|       crm_cd_desc|       mocodes|vict_age|vict_sex|vict_descent|pre
mis_cd|       premis_desc|weapon_used_cd|       weapon_desc|status| status_desc|crm_cd_
1|crm_cd_2|crm_cd_3|crm_cd_4|       location|cross_street|       lat|       lon|
+---+-----+-----+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+-----+-----+-----+
| 1|010304468|2020-01-08 00:00:...|2020-01-08 00:00:...|22:30:00| 03| Southwest| 03
77| 2| 624|BATTERY - SIMPLE ...| 0444 0913| 36| F| B|
501|SINGLE FAMILY DWE...| 400|STRONG-ARM (HANDS...| A0| Adult Other| 624|
|  |  |  |1100 W 39TH|  |  |  |34.0141|-118.2978|
| 2| 0817|2020-09-20 00:00:...|2020-09-19 00:00:...|17:00:00| 17|Devonshire| 17
77| 1| 510| VEHICLE - STOLEN|  |  |  |0|  |  |
101|  |  |  |STREET|  |  |  |IC| Invest Cont| 510|
|  |  |  |9100 RUBIO|  |  |  |34.2367|-118.4955|
| 3|190101086|2020-01-02 00:00:...|2020-01-01 00:00:...|03:30:00| 01| Central| 01
63| 2| 624|BATTERY - SIMPLE ...| 0416 1822 1414| 25| M| H|
102|  |  |  |SIDEWALK| 500|UNKNOWN WEAPON/OT...| IC| Invest Cont| 624|
|  |  |  |700 S HILL|  |  |  |34.0459|-118.2545|
| 4|190101087|2020-01-02 00:00:...|2020-01-01 00:00:...|05:10:00| 01| Central| 01
56| 2| 626|INTIMATE PARTNER ...|1414 1218 2000 18...| 53| F| B|
502|MULTI-UNIT DWELLI...| 400|STRONG-ARM (HANDS...| AA|Adult Arrest| 626|
|  |  |  |300 E 5TH|  |  |  |34.0447|-118.2452|
| 5|190326475|2020-03-01 00:00:...|2020-03-01 00:00:...|21:30:00| 07| Wilshire| 07
84| 1| 510| VEHICLE - STOLEN|  |  |  |0| M| 0|
101|  |  |  |STREET|  |  |  |AA|Adult Arrest| 510|
998|  |  |  |1900 S LONGWOOD ...|  |  |  |34.0375|-118.3506|
+---+-----+-----+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+-----+-----+-----+
+---+-----+-----+-----+-----+-----+-----+-----+
```

Рисунок 2.1 — Импортированный набор данных

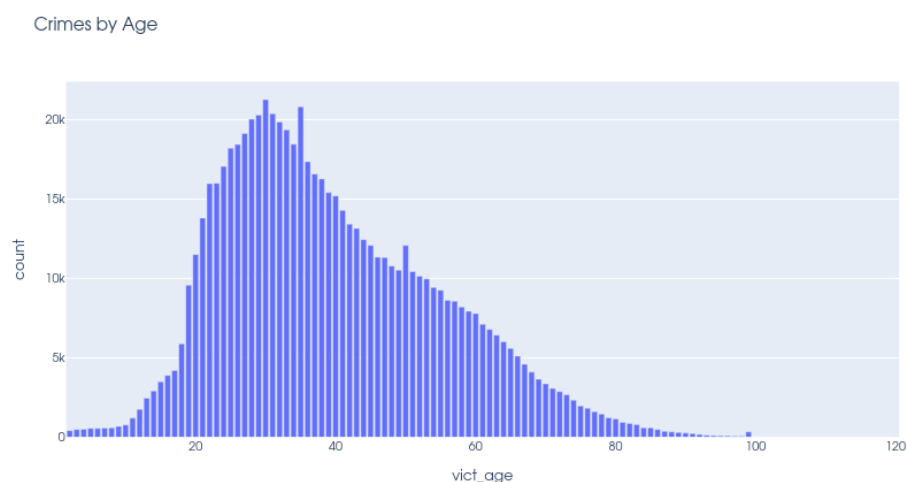
Теперь, когда набор данных находится в Spark, начнем анализ. Найдем основные характеристики жертв преступлений такие, как возраст, пол и происхождение.

Код для составления портрета жертвы приведен в Листинге 2.2.

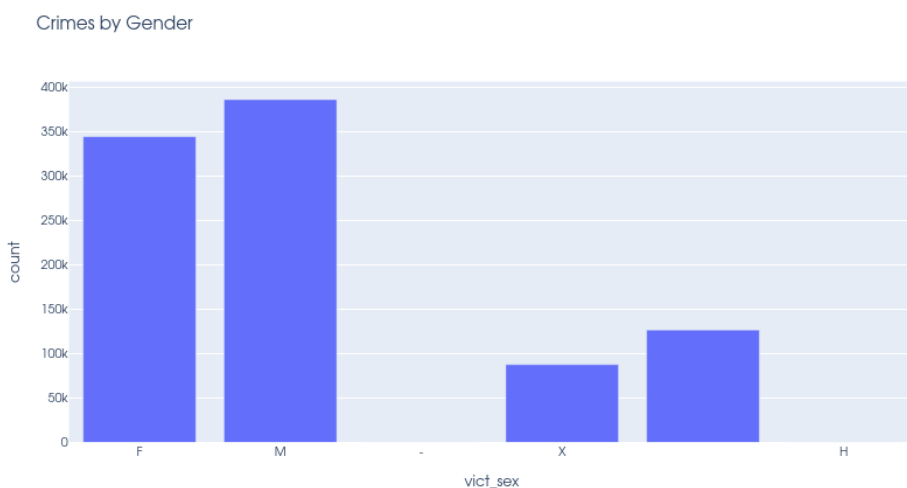
*Листинг 0.2 — Код для составления портрета жертвы*

```
age_groups =  
crime_data.groupBy("vict_age").count().orderBy("vict_age").filter(col("vict_age") > 0)  
fig = px.bar(age_groups, x='vict_age', y='count', title='Crimes by Age')  
fig.show()  
gender_groups = crime_data.groupBy("vict_sex").count()  
fig = px.bar(gender_groups, x='vict_sex', y='count', title='Crimes by Gender')  
fig.show()  
ethnicity_groups = crime_data.groupBy("vict_descent").count()  
fig = px.bar(ethnicity_groups, x='vict_descent', y='count', title='Crimes by Descent')  
fig.show()
```

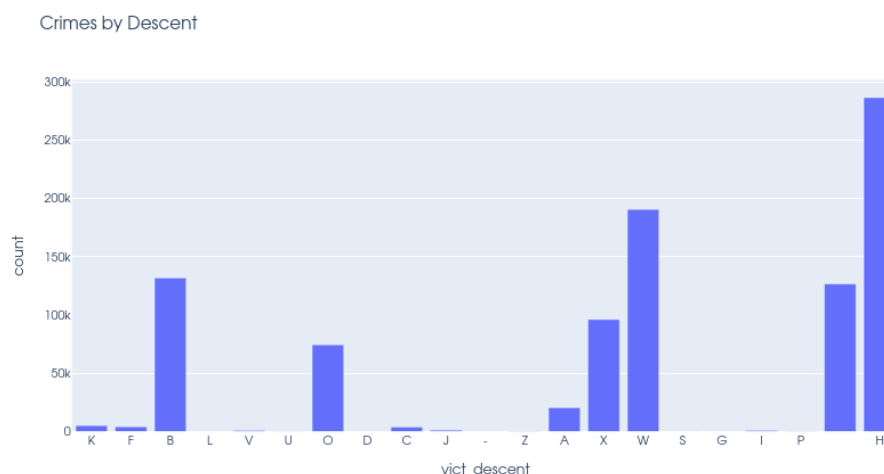
На Рисунках 2.2-2.4 представлены графики зависимости количества преступлений от возраста жертвы, пола и происхождения соответственно.



**Рисунок 2.2 — График зависимости количества преступлений от возраста жертвы**



**Рисунок 2.3 — График зависимости количества преступлений от пола жертвы**



**Рисунок 2.4 — График зависимости количества преступлений от происхождений жертвы**

Согласно полученным данным, наиболее уязвимой группой являются мужчины в возрасте от 25 до 34 лет. Это может быть связано с их более активным образом жизни и повышенной вовлеченностью в различные социальные активности. По этнической принадлежности представители латиноамериканского сообщества составляют наибольшую долю жертв преступлений. Этот факт может быть обусловлен социально-экономическими факторами, такими как уровень доходов, образование, занятость и т.д [2.1].

Теперь перейдем к анализу типов преступлений. Рассмотрим их процентное распределение. Данный анализ позволит выявить ключевые направления криминальной активности, на которых следует сосредоточить усилия правоохранительных органов.

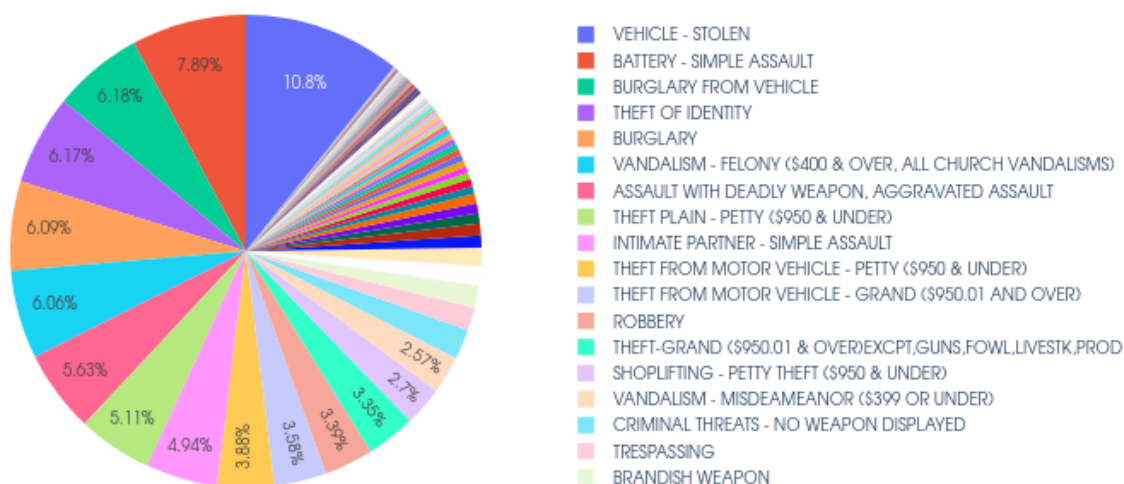
Код для расчета приведен в Листинге 2.3.

*Листинг 0.3 — Код для расчета наиболее характерных мест преступлений*

```
crime_types = crime_data.groupby("crm_cd_desc").count().orderBy(F.col("count").desc())
fig = px.pie(crime_types, values='count', names='crm_cd_desc', title='Main Crime Types')
fig.update_traces(textposition='inside')
fig.update_layout(uniformtext_minsize=12, uniformtext_mode='hide')
fig.show()
```

На Рисунке 2.5 представлен график процентного соотношения типов преступлений.

## Main Crime Types



**Рисунок 2.5 — График процентного соотношения типов преступлений**

Анализируя представленную круговую диаграмму, наглядно отображающую распределение различных видов преступлений, можно выделить ключевые категории, доминирующие в общей криминальной статистике.

Согласно полученным данным, наиболее распространенными типами правонарушений являются угон автотранспорта (10,8%), нападения на граждан (7,89%), кражи из автотранспорта (6,18%), хищение личных данных (6,17%) и кражи со взломом (6,09%). В совокупности эти пять категорий составляют около трети от всех зарегистрированных преступлений. Данные правонарушения наносят ущерб имуществу граждан и организаций, нарушают общественный порядок.

Полученные результаты наглядно свидетельствуют в том, что приоритетными направлениями противодействия преступности должны стать:

- усиление мер по предотвращению угонов автотранспорта, квартирных и уличных краж;
- повышение эффективности защиты личных данных граждан от мошеннических посягательств;
- совершенствование системы охраны общественного порядка [2.2].

Анализ представленной структуры преступности наглядно демонстрирует



необходимость более детального изучения криминальной ситуации в различных районах Лос-Анджелеса. Сравнение показателей по отдельным территориям позволит выявить наиболее проблемные, криминально-опасные зоны, требующие приоритетного внимания правоохранительных органов [2.3].

Код для сравнения уровня преступности в различных районах Лос-Анджелеса для выявления криминально-опасных зон приведен в Приложении А.

На Рисунке 2.6 представлен график распределения преступлений по полицейским участкам.

Map of Crime Counts by District

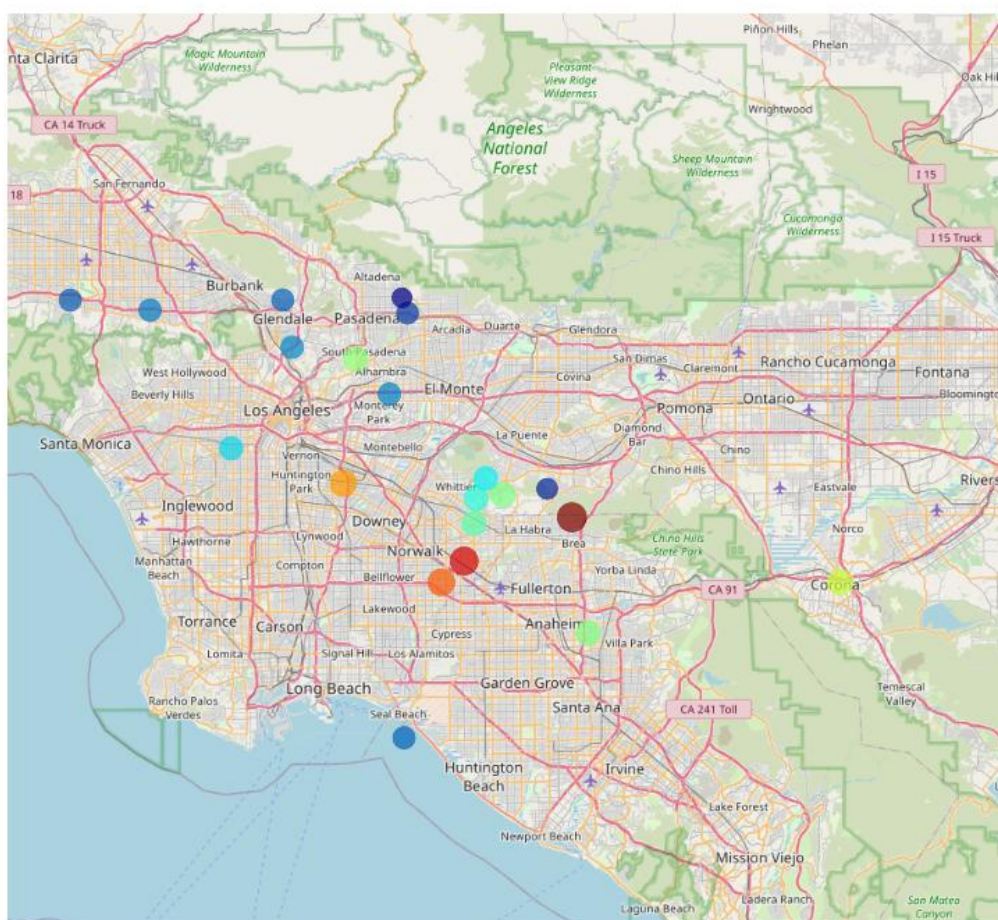


Рисунок 2.6 — График распределения преступлений по полицейским участкам

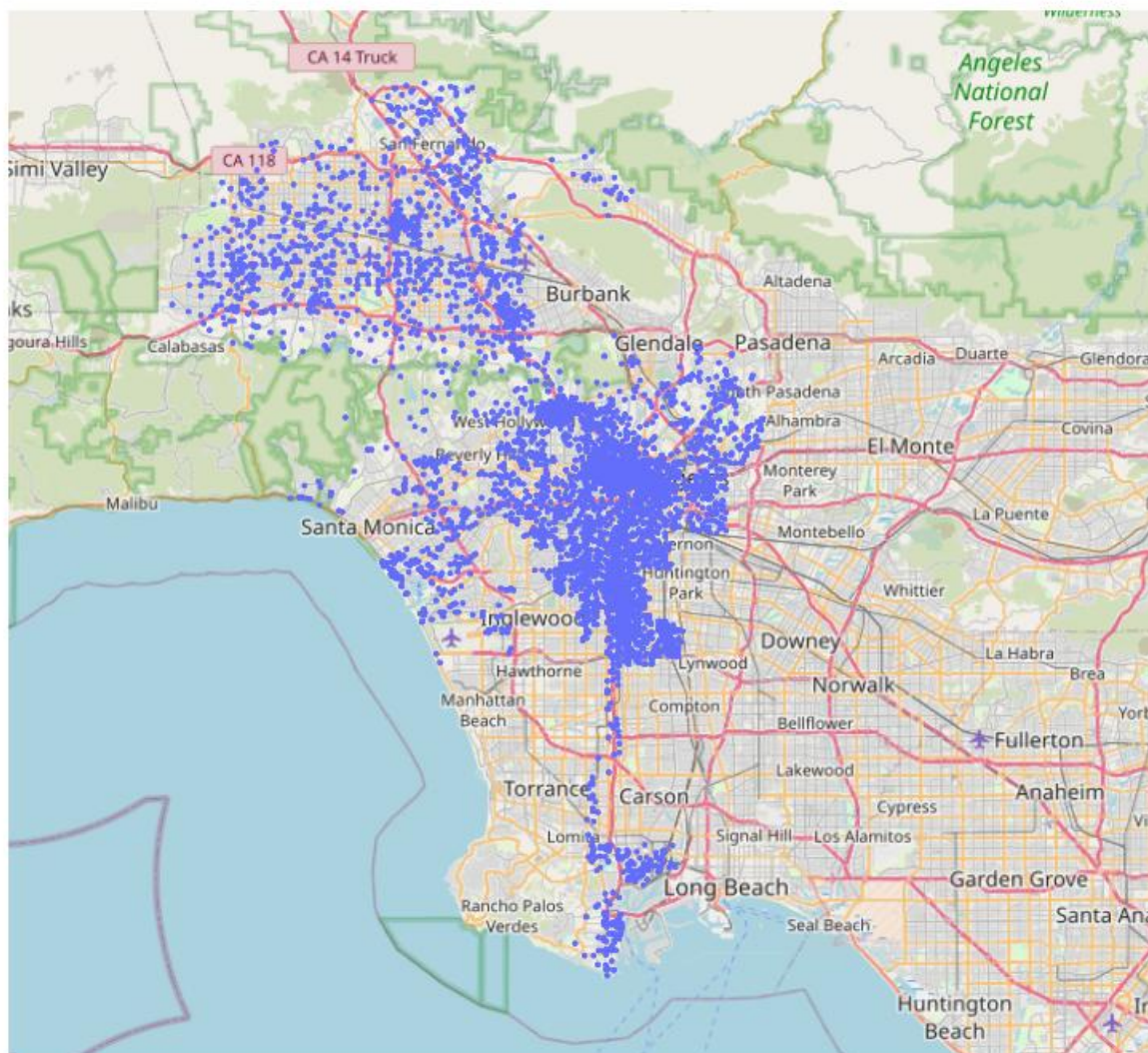
Из графика видно, что основная часть преступлений происходит в Центральном районе (Downtown) и в Южном (South Central) с полицейскими участками Central и 77<sup>th</sup> Street соответственно.

Теперь выделим наиболее опасные преступления, а именно убийство и изнасилование и посмотрим их распределение.



На Рисунке 2.7 представлен график распределения особо опасных преступлений по карте Лос-Анджелеса.

Map of Serious Crimes



**Рисунок 2.7 — График распределения особо опасных преступлений**

График наглядно демонстрирует, что наиболее криминогенными районами города являются Центральный (Downtown) и Южный (South Central) регионы. Данные районы характеризуются высокой концентрацией тяжких уголовных деяний, таких как убийства и изнасилования, что свидетельствует об острой необходимости принятия эффективных мер по обеспечению правопорядка и общественной безопасности в этих районах.

Таким образом, результаты пространственного анализа криминогенной обстановки в Лос-Анджелесе четко указывают на необходимость усиления

профилактической и оперативно-розыскной работы правоохранительных органов в данных районах, что позволит добиться значительного снижения уровня преступности и повышения общественной безопасности.

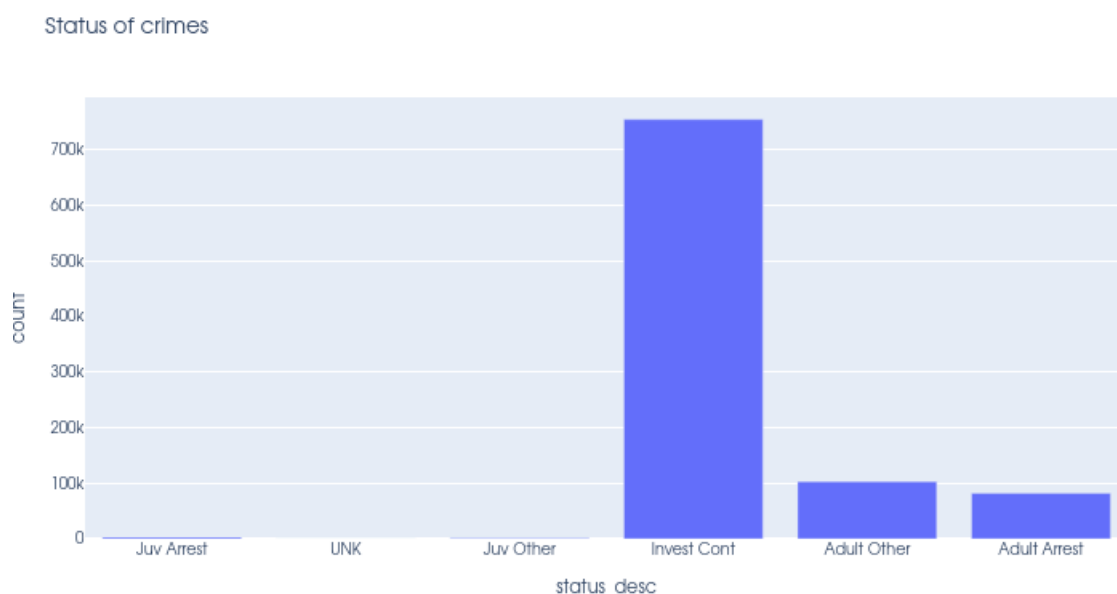
Теперь посмотрим на эффективность работы правоохранительных органов. Выведем количество преступлений по их статусам.

Код для расчета приведен в Листинге 2.4.

*Листинг 0.4 — Код для расчета количества преступлений по их статусу*

```
status_crime = crime_data.groupby("status_desc").count()
fig = px.bar(status_crime, x='status_desc', y='count', title='Status of crimes')
fig.show()
```

На Рисунке 2.8 представлен график соотношения статусов преступлений.



**Рисунок 2.8 — График соотношения статусов преступлений**

Из графика видно, что лишь около 20% преступлений в Лос-Анджелесе находят свое разрешение, в то время как все остальные еще расследуются. Это свидетельствует о серьезных проблемах в работе местных правоохранительных органов [2.4].

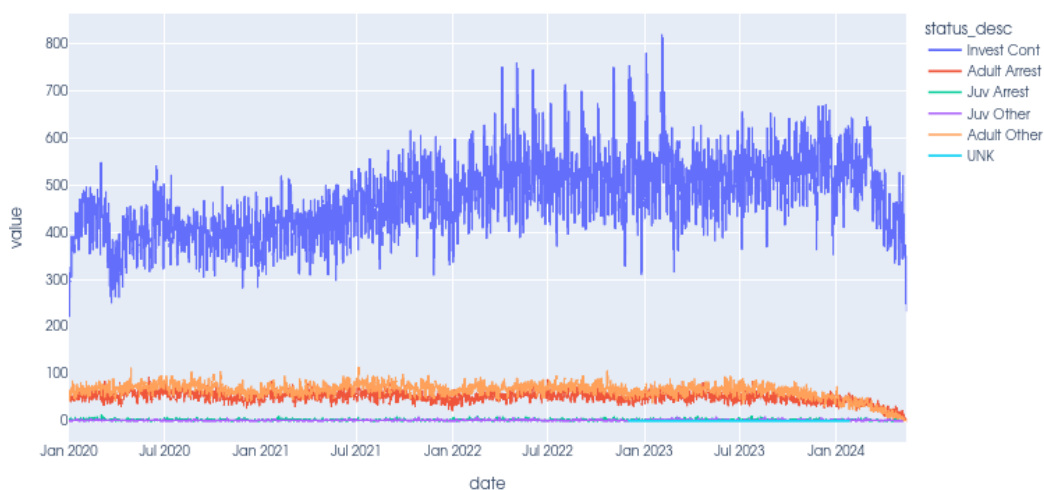
Теперь сгруппируем данные по статусу и времени и оценим эффективность работы полиции, а также посмотрим на динамику раскрытия преступлений в Лос-Анджелесе.

Код для расчета приведен в Листинге 2.5.

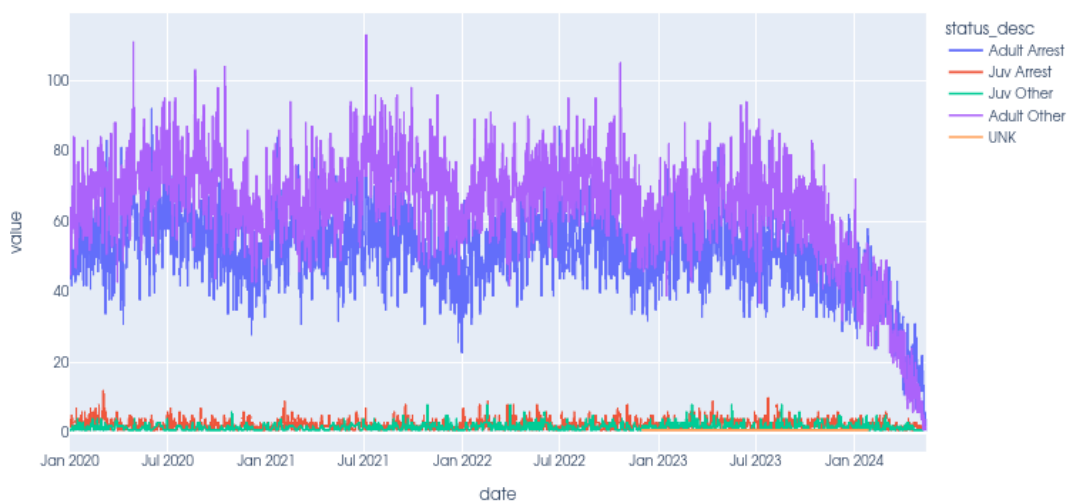
*Листинг 0.5 — Код для оценки эффективности работы полиции*

```
time_series_status = crime_data.withColumn("date", F.to_date("date_rptd")).groupBy("date",
"status_desc").agg(F.count("status_desc").alias("value")).orderBy(F.col("date"))
fig = px.line(time_series_status, x="date", y="value", color="status_desc")
fig.show()
fig = px.line(time_series_status.filter(F.col('status_desc') != 'Invest Cont'), x="date", y="value",
color="status_desc")
fig.show()
```

На Рисунках 2.9-2.10 представлены графики динамики статусов преступлений.



**Рисунок 2.9 — График динамики статусов преступлений**



**Рисунок 2.10 — График динамики статусов раскрытых преступлений**

Из графиков видно, что количество нераскрытых дел растет, пока количество раскрытых держится примерно на одном уровне. Таким образом, несмотря на стабильную долю раскрытых преступлений, наблюдается негативная тенденция к увеличению количества нераскрытых дел. Это указывает на необходимость более пристального внимания к методам расследования и перераспределения ресурсов правоохранительных органов для повышения эффективности их работы [2.5].

Рассмотрим общую динамику преступлений по годам, месяца, дням недели и часам. Это позволит выявить закономерности и тенденции в совершении преступлений. Анализ динамики по различным временным параметрам поможет определить, существуют ли сезонные, недельные или суточные колебания в уровне преступности.

Код для визуализации динамики преступности во времени (по годам, месяцам, дням недели) приведен в Приложении Б.

На Рисунке 2.11 представлены графики общей динамики преступлений по различным временным характеристикам.



**Рисунок 2.11 — Графики общей динамики преступлений по различным временным характеристикам**

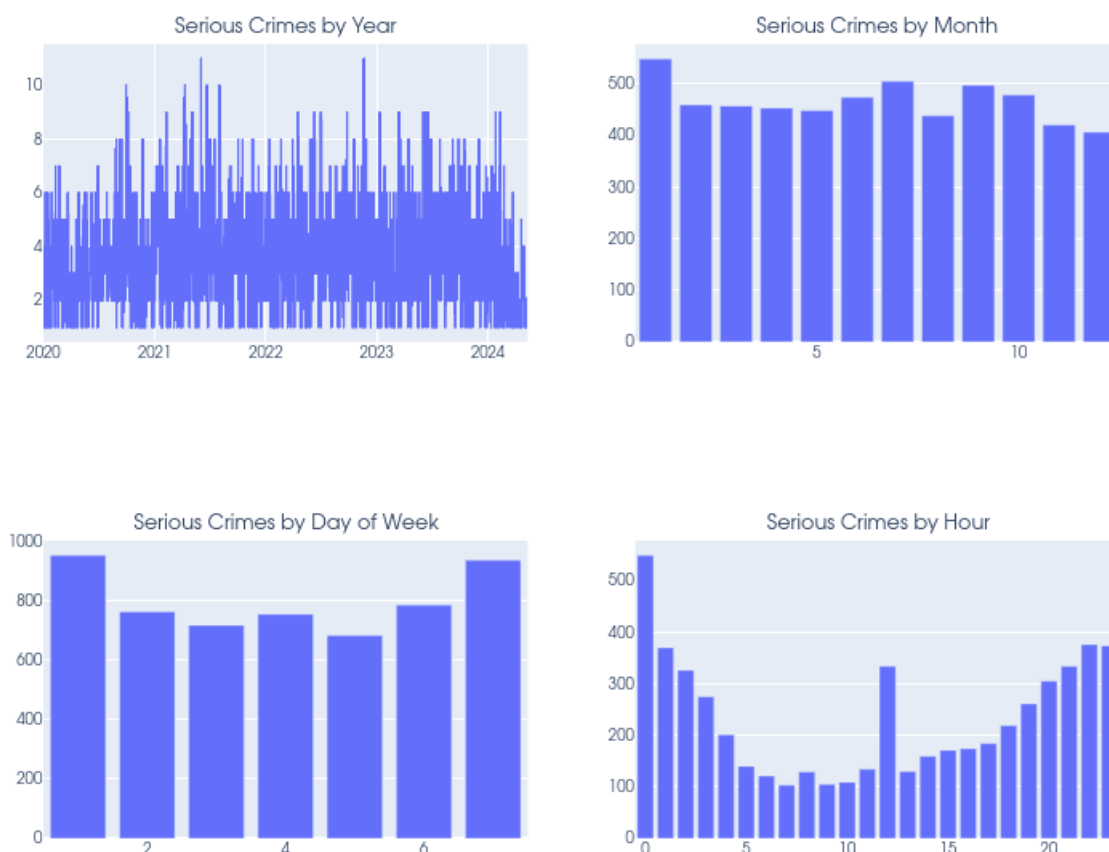
Анализируя графики, можно сделать следующие выводы:

1. Динамика преступлений по годам показывает общую тенденцию к росту уровня преступности в городе за последние 4 года.
2. Помесячная динамика выявляет сезонные колебания преступности. Наибольшее число преступлений совершается в зимние месяцы. Минимальные показатели наблюдаются в осенние и летние месяцы.
3. Распределение преступлений по дням недели демонстрирует, что наибольшее число противоправных деяний приходится на выходные дни (суббота, воскресенье). Это может объясняться более активным времяпрепровождением и досугом граждан в эти дни.
4. Анализ динамики преступлений по часам суток выявляет, что максимальное количество преступлений совершается в вечернее и время (с 18:00 до 20:00 часов). Данная закономерность обусловлена снижением контроля и активности граждан в этот период, а также влиянием факторов алкогольного опьянения.

Полученные данные позволяют более эффективно организовывать работу правоохранительных органов, перераспределяя силы и средства в наиболее криминогенные периоды времени. Кроме того, выявленные тенденции могут использоваться при разработке комплексных профилактических мероприятий, направленных на предупреждение преступности [2.6].

Теперь рассмотрим динамику особо опасных преступлений по годам, месяца, дням недели и часам.

На Рисунке 2.12 представлены графики динамики особо опасных преступлений по различным временным характеристикам.



**Рисунок 2.12 — Графики динамики особо опасных преступлений по различным временным характеристикам**

Анализируя графики динамики особо опасных преступлений (убийства, изнасилования), можно сделать следующие выводы в сравнении с общей динамикой преступлений:

1. Динамика особо опасных преступлений по годам не показывает тенденцию к росту уровня таких преступлений в городе за последние 4 года. Это контрастирует с общей тенденцией снижения преступности в целом.
2. Помесячная динамика особо опасных преступлений выявляет несколько иные сезонные колебания по сравнению с общей картиной. Наибольшее число тяжких преступлений совершается в январе, августе и сентябре, в то время как для общей преступности пик наблюдался только в зимние месяцы.
3. Распределение особо опасных преступлений по дням недели показывает, что наибольшее число таких деяний приходится на

воскресенье и понедельник.

4. Анализ динамики особо опасных преступлений по часам суток выявляет, что максимальное количество таких преступлений совершается несколько позже, чем для общей картины преступности — в интервале с 20:00 до 3:00 часов, достигая пика в 0:00 часов.

Таким образом, динамика тяжких и особо тяжких преступлений имеет некоторые отличия от общей динамики преступности в регионе, требующие отдельного внимания и корректировки профилактических мер правоохранительными органами. Выявленные закономерности помогут более точно прогнозировать криминогенную ситуацию и оптимизировать распределение сил и средств для борьбы с наиболее опасными видами преступлений.

## ЗАКЛЮЧЕНИЕ

В заключение важно отметить, что современное общество сталкивается с серьезной проблемой преступности, которая требует комплексного подхода и анализа данных для разработки эффективных мер по ее предотвращению. Анализ данных является неотъемлемой частью многих сфер деятельности, включая правоохранительную деятельность, и позволяет выявлять тенденции, факторы риска и разрабатывать обоснованные решения.

Представленный в данной курсовой работе анализ преступности на примере Лос-Анджелеса с использованием статистических методов и инструментов демонстрирует практическую значимость подобных исследований. Полученные результаты могут быть использованы для совершенствования подходов к предотвращению преступности и повышению общественной безопасности.

Более того, рассмотренные на примере Лос-Анджелеса инструменты и методы могут быть применены для анализа данных в любом другом городе или регионе, способствуя выработке эффективных стратегий противодействия преступности и укреплению правопорядка.

Цель данной курсовой работы — проанализировать имеющуюся выборку и проанализировать основные характеристики жертв преступлений, основные виды преступлений, сравнить уровень преступности в разных районах Лос-Анджелеса, оценить эффективность работы правоохранительных органов и оценить динамику преступности во времени — достигнута.

В ходе работы были выполнены следующие задачи:

- составлен конвейер для сбора и передачи данных;
- определены ключевые вопросы для проведения анализа;
- проведен анализ в соответствии с целью;
- визуализированы полученные результаты.



# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

## СОЗДАНИЕ КОНВЕЙЕРА ДЛЯ ПЕРЕДАЧИ ДАННЫХ

- 1.1. Crime Data from 2020 to Present [Электронный ресурс]. — Режим доступа: [https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about\\_data](https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/about_data) — Дата доступа: 20.05.2024
- 1.2. Codernet. Что такое MariaDB [Электронный ресурс]. — Режим доступа:  
[https://codernet.ru/articles/sql/chto\\_takoe\\_mariadb\\_gde\\_ispolzuetsya\\_eta\\_sistema\\_upravleniya/](https://codernet.ru/articles/sql/chto_takoe_mariadb_gde_ispolzuetsya_eta_sistema_upravleniya/) — Дата доступа: 20.05.2024
- 1.3. Школа больших данных. Apache Hive [Электронный ресурс]. — Режим доступа: <https://bigdataschool.ru/wiki/hive> — Дата доступа: 20.05.2024
- 1.4. User Guide Apache Sqoop [Электронный ресурс]. — Режим доступа: <https://docs.cloudera.com/sqoop/1.4.7.7.1.6.0/user-guide/index.html> — Дата доступа: 20.05.2024
- 1.5. YandexCloud. Apache Kafka где применяется [Электронный ресурс]. — Режим доступа: <https://cloud.yandex.ru/blog/posts/2021/02/managed-kafka-overview> — Дата доступа: 20.05.2024
- 1.6. Habr. Flume — управление потоками. [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/companies/dca/articles/280386/> — Дата доступа: 20.05.2024
- 1.7. Курс лекций Samsung Innovation Campus [Электронный ресурс]. — Режим доступа:  
[https://myitschool.ru/edu/mod/scorm/player.php?a=2&currentorg=rus\\_sic\\_big\\_data\\_chapter1\\_f\\_final\\_%28s\\_pravkami\\_teksta\\_pod\\_slaidami%29\\_organization&scoid=18](https://myitschool.ru/edu/mod/scorm/player.php?a=2&currentorg=rus_sic_big_data_chapter1_f_final_%28s_pravkami_teksta_pod_slaidami%29_organization&scoid=18) — Дата доступа: 20.05.2024

## АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ

- 2.1. "Криминология: Учебник для вузов" под ред. В.Н. Кудрявцева, В.Е. Эминова. – 4-е изд., перераб. и доп. – М.: Норма, 2016. – 800 с. – ISBN 978-5-91768-679-2
- 2.2. Статья "Анализ преступности с использованием статистических методов" в журнале "Социологические исследования" за 2020 год, № 3, стр. 54-62
- 2.3. Официальный сайт Федеральной службы государственной статистики РФ, раздел "Преступность и правонарушения" [Электронный ресурс]. – Режим доступа: <https://rosstat.gov.ru/folder/13306> — Дата доступа: 20.05.2024
- 2.4. Монография "Пространственный анализ преступности: теория, методология, практика" под ред. Г.Н. Горшенкова. – Н. Новгород: Изд-во ННГУ, 2018. – 372 с. – ISBN 978-5-91326-463-8
- 2.5. Статья "Применение методов интеллектуального анализа данных для выявления закономерностей в преступности" в журнале "Прикладная информатика" за 2019 год, Том 14, № 1, стр. 40-51
- 2.6. Официальный сайт Министерства внутренних дел Российской Федерации, раздел "Статистическая информация" [Электронный ресурс]. – Режим доступа: <https://мвд.рф/Deljatelnost/statistics> — Дата доступа: 20.05.2024

## **ПРИЛОЖЕНИЯ**

Приложение А — Код для сравнения уровня преступности в различных районах Лос-Анджелеса для выявления криминально-опасных зон.

Приложение Б — Код для визуализации динамики преступности во времени (по годам, месяцам, дням недели).

## Приложение А

Код для сравнения уровня преступности в различных районах Лос-Анджелеса для выявления криминально-опасных зон.

*Листинг А.1 — Код для сравнения уровня преступности в различных районах*

```
district_crime_counts = crime_data.groupby('area_name').agg(
    F.avg('lat').alias('latitude'), F.avg('lon').alias('longitude'), F.count('*').alias('counts'))
fig = px.scatter_mapbox(district_crime_counts,
    lat="latitude",
    lon="longitude",
    size="counts",
    color="counts",
    hover_name="area_name",
    color_continuous_scale="jet",
    hover_data=["counts", "latitude", "longitude"],
    zoom=9,
    height=1000, width=1200,
    title="Map of Crime Counts by District")
fig.update_layout(mapbox_style="open-street-map")
fig.show()
serious_crimes = ['110', '113', '121', '122']
serious_crimes_data = crime_data.filter(F.col('crm_cd').isin(serious_crimes))
district_crime_counts = serious_crimes_data
fig = px.scatter_mapbox(district_crime_counts,
    lat="lat",
    lon="lon",
    color_continuous_scale="jet",
    hover_data=["lat", "lon"],
    zoom=9,
    height=1000, width=1200,
    title="Map of Serious Crimes")
fig.update_layout(mapbox_style="open-street-map")
fig.show()
```

## Приложение Б

Код для визуализации динамики преступности во времени (по годам, месяцам, дням недели).

*Листинг Б.1 — Код для визуализации динамики преступности во времени*

```
from pyspark.sql.functions import month, dayofweek, col, count, hour
from plotly.subplots import make_subplots

crime_data = crime_data.withColumn("date", F.to_date("date_rptd"))
crime_data = crime_data.withColumn("Month", month(col("date_occ")))
crime_data = crime_data.withColumn("Day of Week", dayofweek(col("date_occ")))
crime_data = crime_data.withColumn("Hour of Day", hour(col("time_occ")))
crimes_by_year = crime_data.groupBy("date").count().orderBy("date")
crimes_by_month = crime_data.groupBy("Month").count().orderBy("Month")
crimes_by_day = crime_data.groupBy("Day of Week").count().orderBy("Day of Week")
crimes_by_hour = crime_data.groupBy("Hour of Day").count().orderBy("Hour of Day")
fig = make_subplots(rows=2, cols=2, subplot_titles=("Crimes by Year", "Crimes by Month",
"Crimes by Day of Week", "Crimes by Hour"))
fig.add_trace(px.line(crimes_by_year, x="date", y="count").data[0], row=1, col=1)
fig.add_trace(px.bar(crimes_by_month, x='Month', y='count').data[0], row=1, col=2)
fig.add_trace(px.bar(crimes_by_day, x='Day of Week', y='count').data[0], row=2, col=1)
fig.add_trace(px.bar(crimes_by_hour, x='Hour of Day', y='count').data[0], row=2, col=2)
fig.update_layout(height=800, width=1000)
fig.show()

serious_crimes_data = serious_crimes_data.withColumn("date", F.to_date("date_rptd"))
serious_crimes_data = serious_crimes_data.withColumn("Month", month(col("date_occ")))
serious_crimes_data = serious_crimes_data.withColumn("Day of Week",
dayofweek(col("date_occ")))
serious_crimes_data = serious_crimes_data.withColumn("Hour of Day", hour(col("time_occ")))
crimes_by_year = serious_crimes_data.groupBy("date").count().orderBy("date")
crimes_by_month = serious_crimes_data.groupBy("Month").count().orderBy("Month")
crimes_by_day = serious_crimes_data.groupBy("Day of Week").count().orderBy("Day of Week")
crimes_by_hour = serious_crimes_data.groupBy("Hour of Day").count().orderBy("Hour of Day")
```

*Продолжение Листинга Б.1*

```
fig = make_subplots(rows=2, cols=2, subplot_titles=("Serious Crimes by Year", "Serious Crimes  
by Month", "Serious Crimes by Day of Week", "Serious Crimes by Hour"))  
fig.add_trace(px.line(crimes_by_year, x="date", y="count").data[0], row=1, col=1)  
fig.add_trace(px.bar(crimes_by_month, x='Month', y='count').data[0], row=1, col=2)  
fig.add_trace(px.bar(crimes_by_day, x='Day of Week', y='count').data[0], row=2, col=1)  
fig.add_trace(px.bar(crimes_by_hour, x='Hour of Day', y='count').data[0], row=2, col=2)  
fig.update_layout(height=800, width=1000)  
fig.show()
```