

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»
КАФЕДРА ІНФОРМАТИКИ ТА ПРОГРАМНОЇ ІНЖЕНЕРІЇ

КУРСОВА РОБОТА

з дисципліни «Аналіз даних в інформаційних системах»

на тему: «Створення сховища даних російських військовослужбовців, що
беруть участь у війні проти України. Частина 2»

Студента 2 курсу ІТ-03 групи

Спеціальності: 121

«Інженерія програмного забезпечення»

Чабана Антона Євгеновича

«ПРИЙНЯВ» з оцінкою

доц. Ліхоузова Т.А. / доц. Олійник Ю.О.

Підпис

Дата

Київ - 2022 рік

Національний технічний університет України “КПІ ім. Ігоря Сікорського”

Кафедра інформатики та програмної інженерії

Дисципліна Аналіз даних в інформаційно-управляючих системах

Спеціальність 121 "Інженерія програмного забезпечення"

Курс 2 Група ІТ-03

Семестр 4

ЗАВДАННЯ

на курсову роботу студента

Чабана Антона Євгеновича

1.Тема роботи Створення сховища даних російських військовослужбовців, що беруть участь у війні проти України. Частина 2

2.Строк здачі студентом закінченої роботи 19.06.2022

3. Вхідні дані до роботи методичні вказівки до курсової роботи, обрані дані з сайту

4.Зміст розрахунково-пояснювальної записки (перелік питань, які підлягають розробці)

1.Постановка задачі

2.Аналіз предметної області

3.Розробка сховища даних

4.Графічна візуалізація отриманих даних

5.Перелік графічного матеріалу (з точним зазначенням обов’язкових креслень)

6.Дата видачі завдання 16.04.2022

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Назва етапів курсової роботи	Термін виконання етапів роботи	Підписи керівника, студента
1.	Отримання теми курсової роботи	2022-04-16	
2.	Визначення зовнішніх джерел даних	2022-04-28	
3.	Пошук та вивчення літератури з питань курсової роботи	2022-05-01	
4.	Розробка моделі сховища даних	2022-05-16	
4.	Розробка ETL процесів	2022-05-21	
5.	Обґрунтування методів інтелектуального аналізу даних	2022-05-28	
6.	Застосування та порівняння ефективності методів інтелектуального аналізу даних	2022-06-04	
7.	Підготовка пояснювальної записки	2022-06-11	
8.	Здача курсової роботи на перевірку	2022-06-19	
9.	Захист курсової роботи	2022-06-22	

Студент _____
(підпис)

Чабан Антон Євгенович _____
(прізвище, ім'я, по батькові)

Керівник _____
(підпис)

доц. Ліхоузова Т.А _____
(прізвище, ім'я, по батькові)

Керівник _____
(підпис)

доц. Олійник Ю.О. _____
(прізвище, ім'я, по батькові)

"26" червня 2022 р.

АНОТАЦІЯ

Об'єкт дослідження: сховище даних російських військовослужбовців, що беруть участь у війні проти України.

Предмет дослідження: створення сховища даних російських військовослужбовців, що беруть участь у війні проти України та графічне відображення результатів.

Мета роботи: проектування та реалізація сховища даних та ETL процесів.

Дана курсова робота включає в себе: опис проектування, створення та заповнення сховища даних за даною задачею за допомогою фізичної моделі бази даних, опис створення програмного забезпечення для транспортування даних, та графічне відображення.

СХОВИЩЕ ДАНИХ, ФОРМАТУВАННЯ ДАНИХ, ІМПОРТАЦІЯ ДАНИХ, ФІЗИЧНА МОДЕЛЬ БАЗИ ДАНИХ, ETL ПРОЦЕСИ, ВІЗУАЛІЗАЦІЯ ДОСЛІДЖЕННЯ.

ЗМІСТ

ВСТУП	7
1.ПОСТАНОВКА ЗАДАЧІ	8
2.АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	9
3.РОЗРОБКА СХОВИЩА ДАНИХ	10
3.1.РОЗРОБКА МОДЕЛІ СХОВИЩА ДАНИХ	10
3.2.РОЗРОБКА ETL ПРОЦЕСІВ	24
3.3.ЗАВАНТАЖЕННЯ ДАНИХ ЗА ДОПОМОГОЮ ETL ПРОЦЕСІВ....	26
4.ГРАФІЧНА ВІЗУАЛІЗАЦІЯ ОТРИМАНИХ ДАНИХ	27
ВИСНОВКИ	32
ПЕРЕЛІК ПОСИЛАНЬ	33
ДОДАТОК А ТЕКСТИ ПРОГРАМНОГО КОДУ	34

ВСТУП

На сьогоднішній день Україні знаходить в стані війни з росією, щоденно відбуваються військові злочини, мародерство, та вбивства. І поза фронтом активних бойових дій відбувається активна інформаційна війна, щоденно в мережу попадають різні фейки та «вкиди», а завдяки хакерам з різних організацій також трапляється багато різної інформації про військових країни агресора. Як наслідок, нашою командою було вирішено зробити сховище для зберігання даних про всіх військових, чиї дані опинилися в публічному доступі, згрупувати їх та привести до одного загального вигляду, аби надалі була можливість для правосуддя легше їх знайти та покарати.

В рамках даної курсової роботи розроблено сховище даних російських військовослужбовців та їх родин, на основі фізичної моделі бази даних, функціонал якої було розроблено за допомогою SQL скриптів, а саму роботу з базою даних представлено через реалізацію програмного забезпечення для імплементації ETL процесів та інтелектуального аналізу обраних даних.

В ролі системи керування сховищем даних для даної роботи буде виступати PostgreSQL та мова Python для виконання деяких завдань в ході роботи, Power BI та мови PHP для графічної візуалізації проекту.

1. ПОСТАНОВКА ЗАДАЧІ

Під час виконання курсової роботи необхідно виконати наступні завдання:

Створення сховища даних для розміщення всіх даних здобутих у мережі та їх приведення до одного загального виду. Структура курсової роботи узгоджена з керівником. Сховище даних буде містити фінальну таблицю зі згрупованими полями які були обрані з усіх Stage зон. Створення ETL процесів для завантаження даних до сховища, їх отримання зі самого сховища за допомогою запитів, а також оновлення та додавання даних до таблиць вимірів за допомогою SCD першого та другого типів.

Реалізувати спроектоване сховище даних з використанням PostgreSQL 14.2, compiled by Visual C++ build 1914, 64-bit.

Обробити всі файли за допомогою мови Python та привести їх до вигляду CSV файлів, для подальшої імпортації до Stage зони.

Імпортувати файли до бази даних у форматі CSV.

Виконати запити для їх форматування (наприклад дати та звання військових мають бути приведені до єдиного формату).

Виконати ETL процеси для імпортації вже готових даних до фінальної таблиці.

Курсовий проект здати до дедлайну (початок сесії) та виконати у єдиному стилі написання коду (coding style).

2.АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

На сьогоднішній день в мережі з'являється все більше даних російських військовослужбовців. Наша ж задача в тилу використати цю інформацію максимально продуктивно, аби допомогти нашим захисникам у інформаційній війні за Батьківщину. На нашу думку, всі ці данні зможуть бути в подальшому використанні для допомоги кіберполіції та Служби Безпеки України в ідентифікації та знаходженні всіх військових злочинців та подальшого правосуддя. Багато даних містять особисту інформацію, сторінку ВКонтакті, контакти родичів, місце проживання та фотографії що може пришвидшити роботу наших спецслужб. А також допоможе нашій команді виконати безпосередній аналіз даних.

У програмному забезпеченні буде реалізовано наступну функціональність, що включає в себе:

- проектування сховища даних;
- форматування даних до єдиного вигляду
- імпортування даних
- створення ETL процесів для завантаження і оновлення даних;
- створення вибірки даних з сховища;

3.РОЗРОБКА СХОВИЩА ДАНИХ

3.1 Розробка моделі сховища даних

Для виконання курсової роботи було обрано 29 різних джерел даних:

- Телеграм канал CyberWar: <http://t.me/cybwar>
- Телеграм канал InformNapalm: <https://t.me/informnapalm>
- Сайт Генерального Управління розвідкою України:

<https://gur.gov.ua/content/war-criminals-rf.html>

- Сайт організації Anonymous: <https://anonymousleaks.live/>
- Сайт Revenge Monster: <https://leaks.revenge.monster/>
- Платформа Миротворець: <https://myrotvorets.center>

На основі детального опису та проведеного аналізу предметної області було розроблено наступну модель сховища даних російських військових:


























































orc	
 orc_id	integer
 fio /* ПІБ орка */	varchar(100)
 age /* вік */	varchar(20)
 bdate /* дата народження */	varchar(20)
 bplace /* місце народження */	varchar(256)
 nationality /* національність */	varchar(100)
 gender /* (m/f) чоловік / жінка */	varchar(20)
 mstatus /* сімейний стан */	varchar(20)
 address /* повна адреса */	varchar(250)
 reg_address /* адреса реєстрації */	varchar(200)
 tel /* номер телефону */	varchar(20)
 inn /* ІНН */	varchar(20)
 bank /* банківські реквізити */	varchar(512)
 vk /* сторінка VK */	varchar(512)
 sender /* відправник мародер */	varchar(100)
 recipient /* отримувач мародер */	varchar(100)
 tab_num /* Табельний номер */	varchar(20)
 v_ch /* військова частина */	varchar(256)
 branch /* Підрозділ */	varchar(100)
 mil_post /* Військова посада */	varchar(1000)
 nickname /* Позивний */	varchar(100)
 mil_rank /* Військове звання */	varchar(100)
 lnum /* особистий номер */	varchar(20)
 battalion /* батальйон */	varchar(100)
 rota /* рота */	varchar(100)
 mil_ser /* серія військового квитка */	varchar(20)
 mil_num /* номер військового квитка */	varchar(20)
 mil_date /* дата видачі військового квитка */	varchar(20)
 mil_authority /* орган який видав військовий квиток	varchar(100)
 contract_end_date /* дата закінчення контракту */	varchar(20)
 contract_auth /* чий наказ контракту */	varchar(100)
 contract_num /* номер наказу контракту */	varchar(20)
 contract_date	varchar(20)
 edu_lvl /* рівень освіти */	varchar(64)
 study_place /* навчальний заклад */	varchar(1000)
 specialty /* фах */	varchar(100)
 pass_ser /* серія паспорта */	varchar(20)
 pass_num /* номер паспорта */	varchar(20)
 pass_date /* дата видачі паспорта */	varchar(20)
 pass_auth /* орган який видав паспорт */	varchar(128)
 pass_code /* код підрозділу, що видав паспорт */	varchar(20)
 family /* склад сім'ї */	varchar(512)
 children /* кількість дітей */	varchar(20)
 source /* файл джерела даних */	varchar(100)
 photolink /* лінк на файл з фото орка */	varchar(200)
 load_date /* дата завантаження */	timestamp
 gruz200 /* Груз 200 */	boolean
 bucha /* Військові злочини, вбивства, зґвалтування в Київській області	boolean
 description /* Додаткові дані */	text
 mstatus_cl	varchar(100)
 edu_lvl_cl	varchar(100)
 mil_rank_cl	varchar(100)
 bdate_cl	date
 mil_date_cl	date
 pass_date_cl	date
 contract_date_cl	date
 contract_end_date_cl	date

Рисунок 3.1 - Проектування моделі сховища

Для більшої наглядності нашою командою була створена таблиця в якій ми виділили необхідні для імпортації поля (табл. 3.1):

	Name	Table	Field	Description	field type	Sources
	ідентифікатор		orc_id	ідентифікатор	int	
Личная инфа	ФИО		fio	ПІБ орка	varchar(100)	orcs.pdf (табельный номер), 155 обрмп.xlsx, Лютчики РФ.xlsx, Чеченцы.docx, 64омб.xlsx, BBC и ПВО.xlsx, 1327 центр.pdf, Родичі окупантів.xlsx, Полонені РФ ВЧ 12676.xlsx, Полонені РФ ВЧ 31135.xlsxб Полонені РФ ВЧ 31134.txt, АПАКУРТТИ.xls, МОНЧЕГОРСЬК.xls, rf200-1.xlsx, 291 ap.xlsx, fsb.txt, 559 polk.xlsx, 34 бригада.xlsx, 590 zrk.xlsx, 388 зенитно-ракетный полк.xlsx, СДЕК Окупанти-мародери.xlsx, Родственники пленных.xlsx, Список снайперов.docx, СПИСОК гранатометчиков.docx, Сержанты.docx, Список штатный.docx, Список командиров взводов.docx, СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика, Войсковая часть 78018.xlsx, Войсковая часть 77984.xlsx, Боевики ДНР.docx
	Вік	orc	age	вік	varchar(20)	Полонені РФ ВЧ 12676.xlsx, Полонені РФ ВЧ 31135.xlsx
	Дата рождения		bdate	дата народження	varchar(20)	155 обрмп.xlsx, Лютчики РФ.xlsx, Чеченцы.docx, 64омб.xlsx, BBC и ПВО.xlsx, 1327 центр.pdf, Родичі окупантів.xlsx, Полонені РФ ВЧ 12676.xlsx, Полонені РФ ВЧ 31135.xlsx, Полонені РФ ВЧ 31134.txt, rf200-1.xlsx, 291 ap.xlsx, fsb.txt, 559 polk.xlsx, 34 бригада.xlsx, Родственники пленных.xlsx, Войсковая часть 78018.xlsx, Войсковая часть 77984.xlsx, Боевики ДНР.docx
	Место рождения		bplace	місце народження	varchar(100)	155 обрмп.xlsx, Лютчики РФ.xlsx, Родичі окупантів.xlsx, Полонені РФ ВЧ 12676.xlsx, Полонені РФ ВЧ 31135.xlsxб Полонені РФ ВЧ 31134.txt, fsb.txt, Боевики ДНР.docx
	Национальность		nationality	національність	varchar(100)	155 обрмп.xlsx, Лютчики РФ.xlsx, Чеченцы.docx, Родичі окупантів.xlsx, Полонені РФ ВЧ 31134.txt
	Гражданство					Полонені РФ ВЧ 31134.txt

Пол	gender	(m/f) чоловік / жінка	varchar(20)	ВВС и ПВО.xlsx, rf200-1.xlsx
Семейное положение	mstatus	сімейни й стан	varchar(20)	155 обрмп.xlsx, Лютчики РФ.xlsx, ВВС и ПВО.xlsx, Родичі окупантів.xlsx
Край				ВВС и ПВО.xlsx
Город				ВВС и ПВО.xlsx
Адрес фактическо го проживани я	address	повна адреса	varchar(200)	155 обрмп.xlsx, Лютчики РФ.xlsx, Полонені РФ ВЧ 31134.txt, Родственники пленных.xlsx
Данные о регистраци и по месту жительства	reg_address	адреса реєстра ції	varchar(200)	155 обрмп.xlsx, Родичі окупантів.xlsx, fsb.txt
Номер контактного телефона	tel	номер телефо ну	varchar(20)	155 обрмп.xlsx, Родичі окупантів.xlsx, СДЕК Окупанти-мародери.xlsx
ИНН	inn	ИНН	varchar(20)	155 обрмп.xlsx
СНИЛС				155 обрмп.xlsx
Банковские реквизиты	bank	банквс ькі реквізит и	varchar(100)	155 обрмп.xlsx
Страница VK	vk	сторінка VK	varchar(100)	rf200-1.xlsx
Особые отметки				fsb.txt
Удостовере ние личности				Лютчики РФ.xlsx
Дата выдачи удостовере ния				Лютчики РФ.xlsx
Отправите ль	sender	відправ ник мароде р	varchar(100)	СДЕК Окупанти-мародери.xlsx
Получател ь	recipient	отримув ач мароде р	varchar(100)	СДЕК Окупанти-мародери.xlsx
СИГНАТУР А				fsb.txt
Долги				fsb.txt
Авиабилет ы				fsb.txt
Сим-карты				fsb.txt
модемы				fsb.txt

	Автотранспорт				fsb.txt
	Системный				fsb.txt
Военная инфа	Таб.№	tab_num	Табельный номер	varchar(20)	orcs.pdf, 64омб.xlsx, ВВС и ПВО.xlsx, rf200-1.xlsx
	Военная часть	v_ch	військова частина	varchar(100)	Полонені РФ ВЧ 12676.xlsx, Полонені РФ ВЧ 31135.xlsx, Полонені РФ ВЧ 31134.txt, АЛАКУРТТИ.xls, МОНЧЕГОРСЬК.xls, rf200-1.xlsx(раздел персонала), СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика, Войсковая часть 78018.xlsx, Войсковая часть 77984.xlsx
	Раздел персон		Місце служби		orcs.pdf
	РПрс				orcs.pdf, rf200-1.xlsx
	Подразделение	branch	Підрозділ	varchar(100)	Чеченцы.docx, 1327 центр.pdf, Родичі окупантів.xlsx, 559 polk.xlsx, 34 бригада.xlsx, Родственники пленных.xlsx, Список снайперов.docx, СПИСОК гранатометчиков.docx
	Наименование воинской должности	mil_post	Військова посада	varchar(100)	155 обрмп.xlsx, Чеченцы.docx, ВВС и ПВО.xlsx, 1327 центр.pdf, Родичі окупантів.xlsx, Полонені РФ ВЧ 31134.txt, 291 ар.xlsx, 559 polk.xlsx, 34 бригада.xlsx, 590 zrk.xlsx, 388 зенитно-ракетный полк.xlsx, Родственники пленных.xlsx, Сержанты.docx, Список снайперов.docx, СПИСОК гранатометчиков.docx, Список штатный.docx, Список командиров взводов.docx
	Занимаемая воинская должность, штатная категория должности, тарифный разряд		Теж саме, що і військова посада, але в іншому форматі		Льотчики РФ.xlsx
	№ ВУС		Військово-облікова спеціальність		Льотчики РФ.xlsx, ВВС и ПВО.xlsx
	№ ВУС по штату		Військово-облікова спеціальність по штату		155 обрмп.xlsx
	Позывной	nickname	Позивний	varchar(100)	Боевики ДНР.docx

Воинское звание	mil_rank	Військове звання	varchar(100)	155 обрмп.xlsx, Льотчики РФ.xlsx, Чеченцы.docx, 64омб.xlsx, ВВС и ПВО.xlsx, 1327 центр.pdf, Родичі окупантів.xlsx, АЛАКУРТТІ.xls, МОНЧЕГОРСЬК.xls, rf200-1.xlsx, 291 ар.xlsx, 559 polk.xlsx, 34 бригада.xlsx, 590 zrk.xlsx, 388 зенитно-ракетный полк.xlsx, Родственники пленных.xlsx, Список снайперов.docx, СПИСОК гранатометчиков.docx, Сержанты.docx, Список штатный.docx, Список командиров взводов.docx, Войсковая часть 78018.xlsx, Войсковая часть 77984.xlsx, Боевики ДНР.docx
Воинское звание по штату		Теж саме?		Чеченцы.docx, Родичі окупантів.xlsx, 590 zrk.xlsx
Дата присвоения воинского звания, кем когда присвоено № приказа				Льотчики РФ.xlsx
Личный номер	Inum	особистий номер	varchar(20)	orcs.pdf, 155 обрмп.xlsx, 64омб.xlsx, ВВС и ПВО.xlsx, АЛАКУРТТІ.xls, МОНЧЕГОРСЬК.xls, rf200-1.xlsx, 291 ар.xlsx, 559 polk.xlsx, Родственники пленных.xlsx, Войсковая часть 78018.xlsx, Войсковая часть 77984.xlsx
Дата поступления в ВС РФ				155 обрмп.xlsx, Льотчики РФ.xlsx
Военный комиссариат				155 обрмп.xlsx, Льотчики РФ.xlsx
Дата прохождения последнего ВВК				155 обрмп.xlsx
Сведения об участии в боевых действиях				155 обрмп.xlsx
Квалификация				155 обрмп.xlsx (БТГР), Льотчики РФ.xlsx
Полк				ВВС и ПВО.xlsx
БРИГАДА/ДИВИЗИЯ				ВВС и ПВО.xlsx
Объединение				ВВС и ПВО.xlsx
Батальон	battalion	батальон	varchar(100)	590 zrk.xlsx, Список штатный.docx, СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика
Рота	rota	рота	varchar(100)	590 zrk.xlsx, Список штатный.docx, СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика

	Взвод				СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика
	Отделение				СПИСОК_личного_состава_70_мсп_в_ч_71718,_г_Шали,_Чеченская_Республика
	Вид выплаты				АЛАКУРТТИ.xls, МОНЧЕГОРСЬК.xls
	Сумма (руб.)				АЛАКУРТТИ.xls, МОНЧЕГОРСЬК.xls
Военное дополни тельно	Дата принятия д\д				Чеченцы.docx
	ГР_СПЕЦ				ВВС и ПВО.xlsx
	ВГР_СПЕЦ_гр5				ВВС и ПВО.xlsx
	ШДК				ВВС и ПВО.xlsx
	ТР				ВВС и ПВО.xlsx
	ОВУ				ВВС и ПВО.xlsx
	Дата прибытия в в/часть по предписанию				ВВС и ПВО.xlsx
	Фактическая дата прибытия в в/часть				ВВС и ПВО.xlsx
	Зачисление приказ дата				ВВС и ПВО.xlsx
	Зачисление приказ номер				ВВС и ПВО.xlsx
	Зачисление приказ чей				ВВС и ПВО.xlsx
	Дата заведения ТН				ВВС и ПВО.xlsx
	Телефон офицера				ВВС и ПВО.xlsx
	Ответственный исполнитель (в кадровом органе вида, рода, округа)				ВВС и ПВО.xlsx
	Прием дел и должности ДАТА				ВВС и ПВО.xlsx

	Наставник				BBC и ПВО.xlsx
Военный билет	Серия (военный билет)	mil_ser	серія військов ого квитка	varchar(20)	155 обрмп.xlsx
	Номер (военный билет)	mil_num	номер військов ого квитка	varchar(20)	155 обрмп.xlsx
	Дата выдачи (военный билет)	mil_date	дата видачі військов ого квитка	varchar(20)	155 обрмп.xlsx
	Орган выдавший военный билет	mil_authority	орган який видав військов ий квиток	varchar(100)	155 обрмп.xlsx
Контракт	Окончание контракта	contract_end_date	дата закінчен ня контрак ту	varchar(20)	155 обрмп.xlsx (срок действия контракта), Родственники пленных.xlsx
	Чей приказ	contract_auth	чий наказ контрак ту	varchar(100)	155 обрмп.xlsx
	№ приказа	contract_number	номер наказу контрак ту	varchar(20)	155 обрмп.xlsx
	Дата приказа	contract_date	дата наказу контрак ту	varchar(20)	155 обрмп.xlsx, Родственники пленных.xlsx
	Дата начала последнего контракта, на сколько заключен контракт, тип контракта, дата окончания контракта		Теж саме, що і вище, але треба розбити по колонках		Льотчики РФ.xlsx
Образование	Тип образования	edu_lvl	рівень освіти	varchar(20)	155 обрмп.xlsx, BBC и ПВО.xlsx (_ВВУЗ/ВУЦ), Полонені РФ ВЧ 31134.txt
	Образовательное учреждение	study_place	навчальний заклад	varchar(100)	155 обрмп.xlsx (Тип образовательного учреждения), BBC и ПВО.xlsx

	Начальник отделения кадров ВУЗа				ВВС и ПВО.xlsx
	Специальность	specialty	фах	varchar(100)	155 обрмп.xlsx
	Дата окончания образовательного учреждения (по наивысшему уровню образования)				155 обрмп.xlsx, ВВС и ПВО.xlsx
	Диплом с отличием				ВВС и ПВО.xlsx
	Золотая медаль				ВВС и ПВО.xlsx
Паспорт	Серия (паспорт)	pass_ser	серия паспорта	varchar(20)	155 обрмп.xlsx, 64омб.xlsx (серия, номер), rf200-1.xlsx, fsb.txt
	Номер (паспорт)	pass_num	номер паспорта	varchar(20)	155 обрмп.xlsx, Льотчики РФ.xlsx, 64омб.xlsx (паспорта), rf200-1.xlsx, fsb.txt
	Дата выдачи (паспорт)	pass_date	дата выдачи паспорта	varchar(20)	155 обрмп.xlsx, Льотчики РФ.xlsx, 64омб.xlsx, rf200-1.xlsx, fsb.txt
	Орган выдавший паспорт	pass_auth	орган який видав паспорт	varchar(100)	155 обрмп.xlsx, Льотчики РФ.xlsx, 64омб.xlsx, rf200-1.xlsx, fsb.txt
	Код подразделения	pass_code	код підрозділу, що видав паспорт	varchar(20)	155 обрмп.xlsx, rf200-1.xlsx, fsb.txt
	Подтип паспорта				rf200-1.xlsx
	Служебный паспорт				155 обрмп.xlsx
Родственники	Состав семьи	family	склад сім'ї	varchar(100)	155 обрмп.xlsx, Льотчики РФ.xlsx, Родичі окупантів.xlsx, Родственники пленных.xlsx
	Кол-во детей	children	кількість дітей	varchar(20)	ВВС и ПВО.xlsx, Родичі окупантів.xlsx
	Кол-во членов семьи с в/служащим				Льотчики РФ.xlsx
	Серия (паспортные данные)				155 обрмп.xlsx

	родственники)					
	Номер (паспортные данные родственников)					155 обрмп.xlsx
	Дата выдачи (паспортные данные родственников)					155 обрмп.xlsx
	Орган выдавший паспорт (паспортные данные родственников)					155 обрмп.xlsx
	Код подразделения (паспортные данные родственников)					155 обрмп.xlsx
	Банковские реквизиты родственников					155 обрмп.xlsx
Другое	Жилье - служебное					ВВС и ПВО.xlsx
	Жилье - общежитие					ВВС и ПВО.xlsx
	Жилье - приспособленное помещение					ВВС и ПВО.xlsx
	Жилье - поднаем					ВВС и ПВО.xlsx
	Жилье - у родственников					ВВС и ПВО.xlsx
	Получено подъемное пособие					ВВС и ПВО.xlsx
	Обеспечено вещевым имуществом					ВВС и ПВО.xlsx
	Жена трудоустроена					ВВС и ПВО.xlsx
	Обеспечение детей -					ВВС и ПВО.xlsx

	детским садом					
	Обеспечение детей - школой					BBC и ПВО.xlsx
	Медицинское обеспечение членов семей					BBC и ПВО.xlsx
	Примечание					BBC и ПВО.xlsx
	source		source	файл джерела даних	varchar(100)	
	Name	Table	Field	Description	field type	Sources

У моделі сховища за типом зірка (відформатованої до однієї таблиці). В нашій таблиці є виміри дат, сімейних статусів, військових звань та рівня навчання. В наступній таблиці видно детальний опис полів(табл. 3.2).

Таблица 3.2 – Таблица Orc та опис полів

ФИО	fio
Вік	age
Дата рождения	bdate
Место рождения	bplace
Национальность	nationality
Пол	gender
Семейное положение	mstatus
Адрес фактического проживания	address
Данные о регистрации по месту жительства	reg_address
Номер контактного телефона	tel
ИНН	inn
Банковские реквизиты	bank
Страница ВК	vk
Отправитель	sender
Получатель	recipient
Таб.No	tab_num
Военная часть	v_ch

Подразделение	branch
Наименование воинской должности	mil_post
Позывной	nickname
Воинское звание	mil_rank
Личный номер	lnum
Батальон	battalion
Рота	rota
Серия (военный билет)	mil_ser
Номер (военный билет)	mil_num
Дата выдачи (военный билет)	mil_date
Орган выдавший военный билет	mil_auth ority
Окончание контракта	contract _end_da te
Чей приказ	contract _auth
№ приказа	contract _num
Дата приказа	contract _date
Тип образования	edu_lvl
Образовательное учреждение	study_pl ace
Специальность	specialty
Серия (паспорт)	pass_se r
Номер (паспорт)	pass_nu m
Дата выдачи (паспорт)	pass_da te
Орган выдавший паспорт	pass_au th
Код подразделения	pass_co de
Состав семьи	family
Кол-во детей	children
source	source
дата завантаження	load_dat e

На основі спроектованої моделі було створено сховище даних військовослужбовців, за допомогою нижче наведених скриптів використовуючи PostgreSQL.

Створення таблиці Orc в сховищі даних:

```
create table orc
(
    orc_id          serial
        constraint orc_pkey
            primary key,
    fio             varchar(100) not null,
    age             varchar(20),
    bdate           varchar(20),
    bplace          varchar(256),
    nationality      varchar(100),
    gender          varchar(20) default 'm'::character varying,
    mstatus         varchar(20),
    address         varchar(250),
    reg_address     varchar(200),
    tel             varchar(20),
    inn             varchar(20),
    bank            varchar(512),
    vk              varchar(512),
    sender          varchar(100),
    recipient       varchar(100),
    tab_num         varchar(20),
    v_ch            varchar(256),
    branch          varchar(100),
    mil_post        varchar(1000),
    nickname        varchar(100),
    mil_rank        varchar(100),
    lnum            varchar(20),
    battalion       varchar(100),
    rota            varchar(100),
    mil_ser         varchar(20),
    mil_num         varchar(20),
    mil_date        varchar(20),
    mil_authority   varchar(100),
    contract_end_date varchar(20),
    contract_auth   varchar(100),
    contract_num    varchar(20),
    contract_date   varchar(20),
    edu_lvl         varchar(64),
    study_place     varchar(1000),
    specialty       varchar(100),
    pass_ser        varchar(20),
    pass_num        varchar(20),
    pass_date       varchar(20),
    pass_auth       varchar(128),
    pass_code       varchar(20),
    family          varchar(512),
    children        varchar(20),
    source          varchar(100),
    photolink       varchar(200),
    load_date       timestamp default CURRENT_TIMESTAMP,
```

```

    gruz200          boolean      default false,
    bucha            boolean      default false,
    description      text,
    mstatus_cl       varchar(100),
    edu_lvl_cl       varchar(100),
    mil_rank_cl      varchar(100),
    bdate_cl         date,
    mil_date_cl      date,
    pass_date_cl     date,
    contract_date_cl date,
    contract_end_date_cl date
);

comment on column orc.fio is 'ПІВ орка';

comment on column orc.age is 'вік';

comment on column orc.bdate is 'дата народження';

comment on column orc.bplace is 'місце народження';

comment on column orc.nationality is 'національність';

comment on column orc.gender is '(m/f) чоловік / жінка';

comment on column orc.mstatus is 'сімейний стан';

comment on column orc.address is 'повна адреса';

comment on column orc.reg_address is 'адреса реєстрації';

comment on column orc.tel is 'номер телефону';

comment on column orc.inn is 'ІНН';

comment on column orc.bank is 'банківські реквізити';

comment on column orc.vk is 'сторінка VK';

comment on column orc.sender is 'відправник мародер';

comment on column orc.recipient is 'отримувач мародер';

comment on column orc.tab_num is 'Табельний номер';

comment on column orc.v_ch is 'військова частина';

comment on column orc.branch is 'Підрозділ';

comment on column orc.mil_post is 'Військова посада';

comment on column orc.nickname is 'Позивний';

comment on column orc.mil_rank is 'Військове звання';

comment on column orc.lnum is 'особистий номер';

```

```

comment on column orc.battalion is 'батальйон';

comment on column orc.rota is 'рота';

comment on column orc.mil_ser is 'серія військового квитка';

comment on column orc.mil_num is 'номер військового квитка';

comment on column orc.mil_date is 'дата видачі військового квитка';

comment on column orc.mil_authority is 'орган який видав військовий квиток';

comment on column orc.contract_end_date is 'дата закінчення контракту';

comment on column orc.contract_auth is 'чий наказ контракту';

comment on column orc.contract_num is 'номер наказу контракту';

comment on column orc.edu_lvl is 'рівень освіти';

comment on column orc.study_place is 'навчальний заклад';

comment on column orc.specialty is 'фах';

comment on column orc.pass_ser is 'серія паспорта';

comment on column orc.pass_num is 'номер паспорта';

comment on column orc.pass_date is 'дата видачі паспорта';

comment on column orc.pass_auth is 'орган який видав паспорт';

comment on column orc.pass_code is 'код підрозділу, що видав паспорт';

comment on column orc.family is 'склад сімї';

comment on column orc.children is 'кількість дітей';

comment on column orc.source is 'файл джерела даних';

comment on column orc.photolink is 'лінк на файл з фото орка';

comment on column orc.load_date is 'дата завантаження';

comment on column orc.gruz200 is 'Груз 200 ';

comment on column orc.bucha is 'Військові злочини, вбивства, зґвалтування в Київській області ';

comment on column orc.description is 'Додаткові дані';

alter table orc
    owner to pgadmin;

```

Полонені РФ ВЧ 31135.xlsx - Аркуш1 (1)	
п/п	integer
ПІБ	text
ВІК	integer
ДАТА НАРОДЖЕННЯ	text
МІСЦЕ НАРОДЖЕННЯ	text
ВІЙСКОВА ЧАСТИНА	text

3.2 Розробка ETL процесів

Для проектування та подальшої роботи з сховищем даних потрібно врахувати комплекс методів, які реалізують процес переносу початкових даних в аналітичний формат, який буде підтримуватись в сховищі даних та не порушуватиме цілісність системи. Для цього опишемо основні функції ETL процесів, що включають:

Процес завантаження.

Цей етап передбачає зтягування даних довільної якості в ETL для подальшої обробки. Вхідні дані можуть бути в джерелах різних типів і форматів, бути створені в різних додатках, і, крім того, можуть використовувати різне кодування. Саме тому в подальшому для вирішення завдань аналізу дані повинні бути перетворені в єдиний універсальний формат, щоб запобігти вмісту різних факторів, які ускладнюють коректний аналіз даних.

Процес валідації даних.

На цьому етапі головною задачею є перевірка даних на коректність і повноту, після чого складається звіт про помилки для виправлення та узагальнення даних. Первинне очищення даних в процесі ETL носить переважно технічний характер, в якому основне завдання - підготувати дані до завантаження в сховище. Вторинна очистка в аналітичній системі є користувальницькою та спрямована на підготовку даних до вирішення конкретної аналітичної задачі. Завдяки чому обидва етапи очищення однаково важливі і необхідні.

Перетворення даних.

Даний етап ETL процесу потрібний для підготовки даних до розміщення в сховище даних і приведення їх до найбільш зручного виду для подальшого аналізу. При цьому повинна враховуватися вимога рівня якості даних. Тому в процесі перетворення можуть бути задіяні найрізноманітніші інструменти, які можуть бути як і найпростішими засобами ручного редагування даних так і реалізація досить складних методів обробки та очищення даних.

Завантаження даних в сховище.

Після того як дані отримано з різних джерел і виконано перетворення, агрегація і очищення даних, здійснюється останній етап ETL, тобто завантаження даних в сховище. Процес завантаження полягає в перенесенні даних з проміжних таблиць в структури сховища даних. Саме тому, щоб пришвидшити завантаження даних потрібно заздалегідь перевірити дані на повноту та коректність, а після спроектувати та обрати найоптимальніший процес заповнення сховища. При завантаженні даних в сховище переноситься не вся інформація з джерел, а тільки та, яка була змінена протягом проміжного часу, що пройшов з попереднього завантаження.

3.3 Завантаження даних за допомогою ETL процесів

Для заповнення сховища даних російських військовослужбовців було реалізовано наступне:

- 1) Для транспортування даних до Stage зони:
 - 1) Мовою Python було реалізовано функцію яка перетворює файли `xlsx` та `pdf` у `csv`.
 - 2) Засобами `DataGrip` було транспортовано дані з файлів `CSV` до таблиць
 - 3) Засобами `SQL` було здійснено форматування даних до єдиного формату (див. Додатки)
- 2) Було отримано дані з Stage зони для заповнення сховища даних. А саме:
 - 1) За допомогою засобів `SQL` було транспортовано дані відповідних відформатованих

4. ГРАФІЧНА ВІЗУАЛІЗАЦІЯ ОТРИМАНИХ ДАНИХ

Засобами програми Power BI було виконано деякі графіки для наглядності картини отриманої в наслідок збору даних з різних джерел, їх очистки та форматації.

На рисунку 4.1 представлено графік кількості військових залежно від їх віку та військово звання

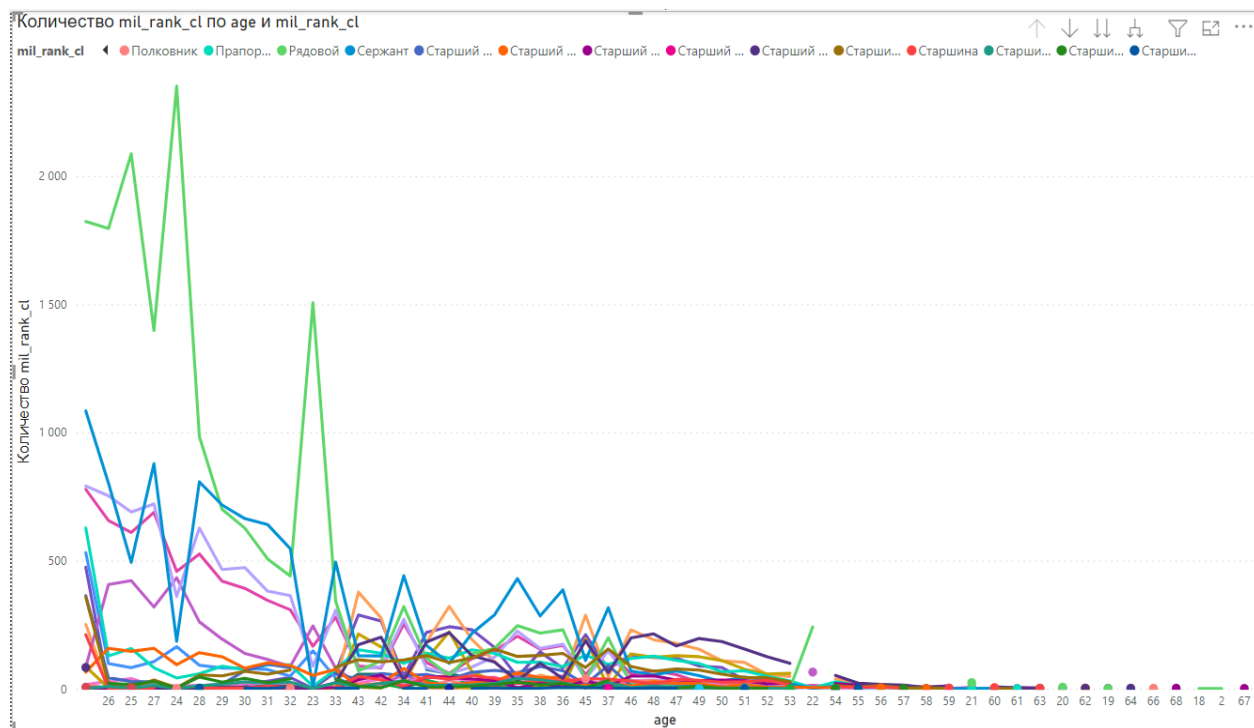


Рис. 4.1 – Кількість звань залежно від віку військовослужбовців

Дивлячись на графік добре видно що чим молодше військовий тим більше серед них молодшого складу. А з віком кількість офіцерського складу зростає.

На рисунку 4.2 представлена загальна кількість рангів та їх поділення на відсотки

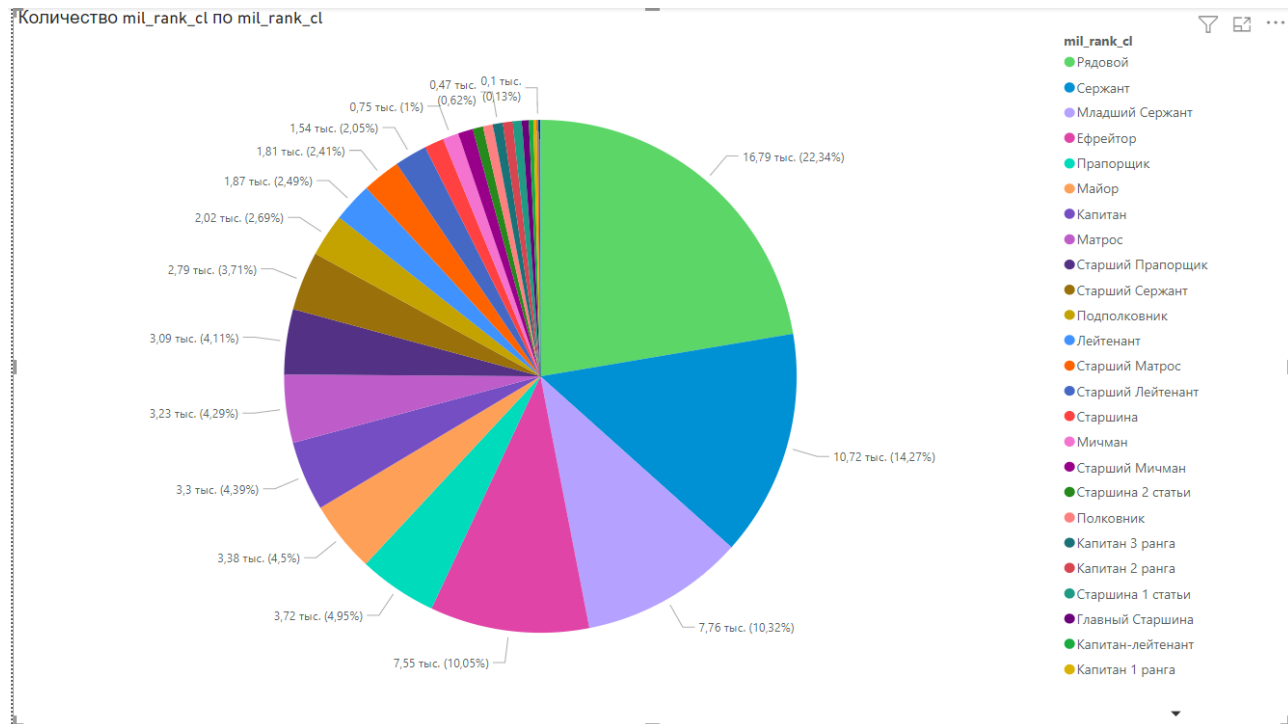


Рис. 4.2 – Загальна кількість рангів та їх доля у відсотках

Дивлячись на цей графік не виконуючи аналіз можна зрозуміти що велика доля військових ймовірно контрактники які пішли на війну «в погоні за грошима» і мають звання молодшого складу.

На рисунку 4.3 представлена загальна картина поділення військових по віку

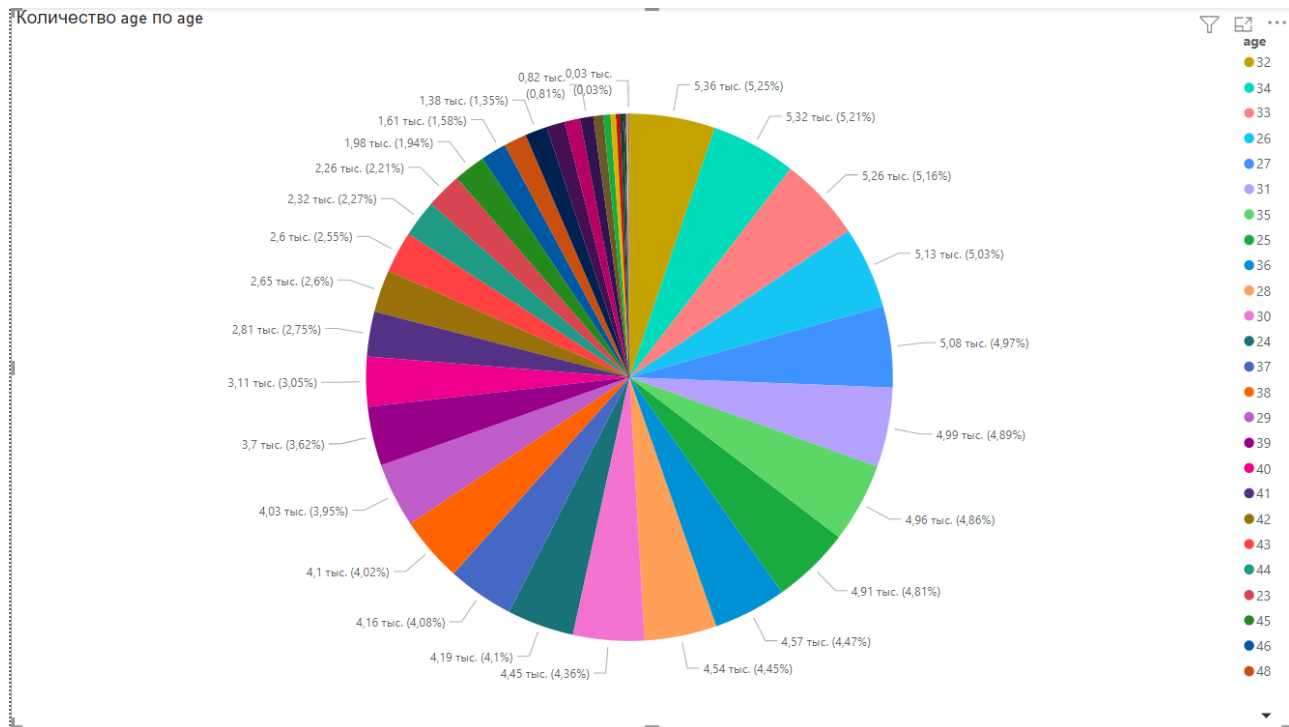


Рис. 4.3 – поділення військових по віку та їх доля у відсотках

На рисунку 4.4 графічно відображено кількість чоловіків та жінок що беруть участь у війні

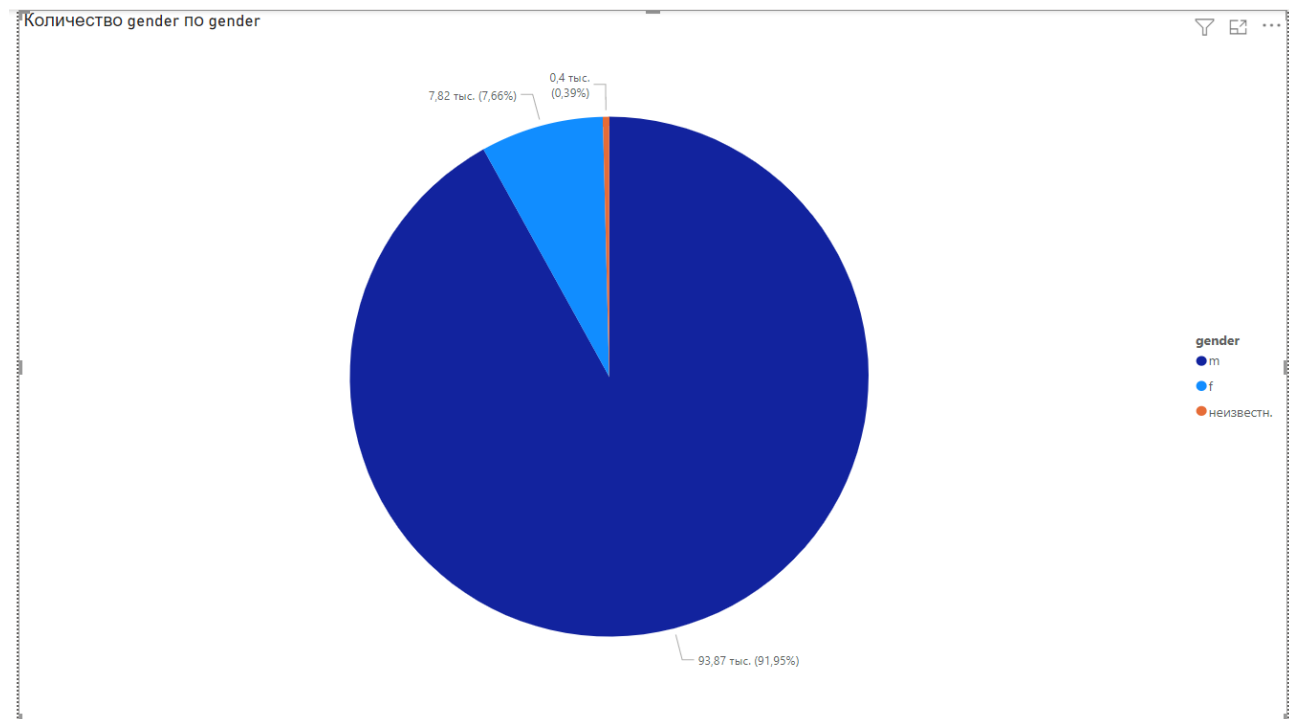


Рис. 4.4 – відсоткове поділення за статтю

На рисунку 4.5 відображено те, який рівень освіти мають російські військовослужбовці

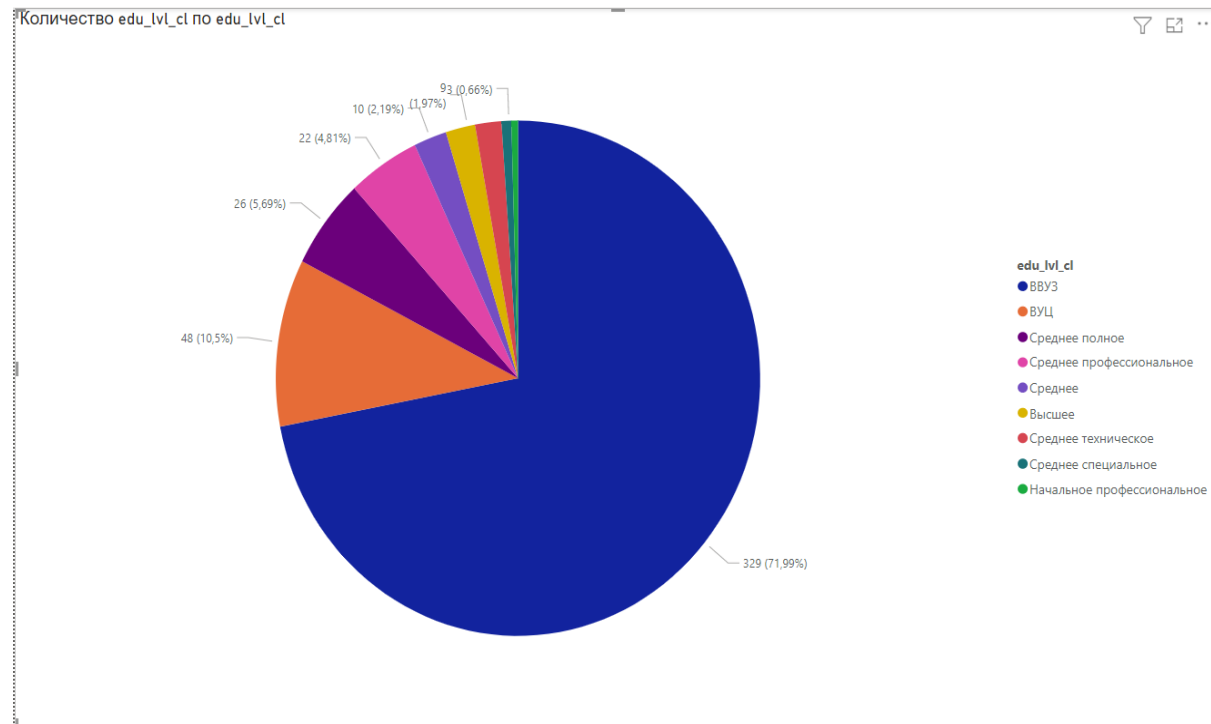


Рис. 4.5 – рівень освіти

Дивлячись на цей графік можна зрозуміти що більшість проаналізованих військових закінчували військові училища різних типів.

На рисунку 4.6 відображено те в якому сімейному статусі знаходяться проаналізовані військовослужбовці.

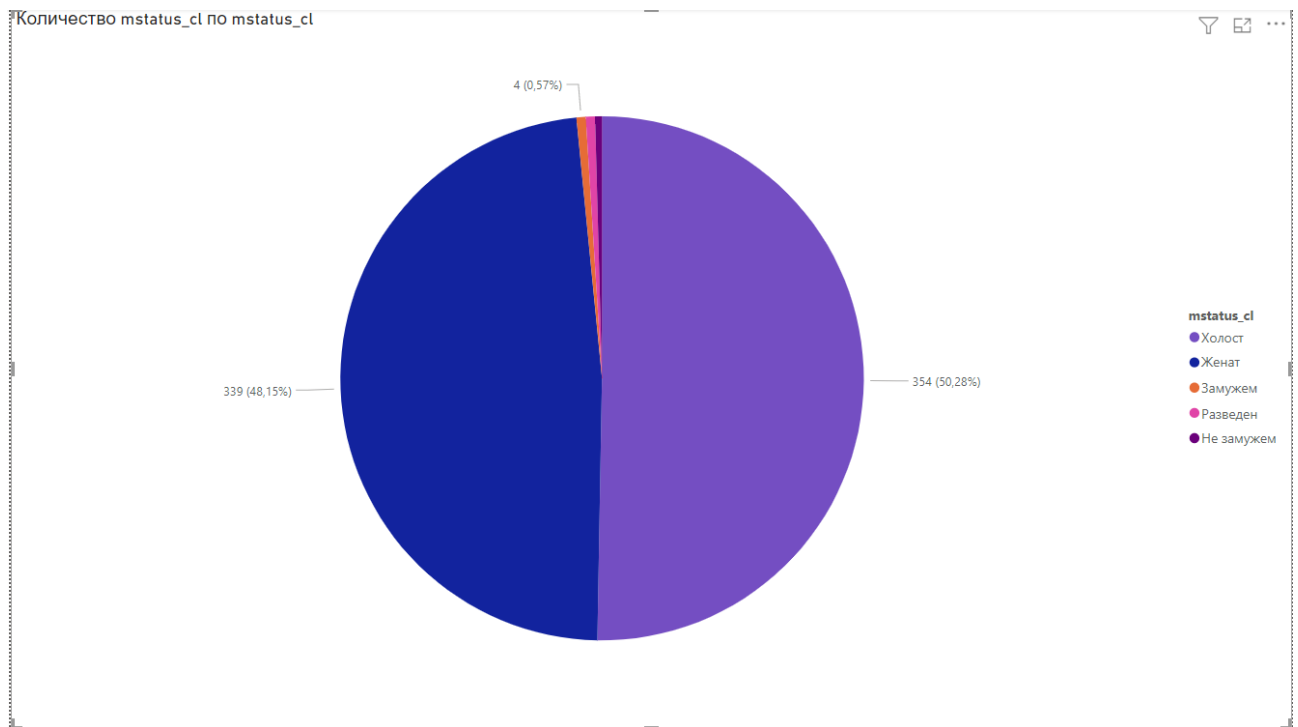


Рис. 4.6 – сімейний статус військовослужбовців

Оглядаючи цей графік добре видно що більше половини військових одинокі і їх може нічого не тягнути додому, що і зумовлює їх жорстокі дії. Проте присутній і великий відсоток сімейних чоловіків.

ВИСНОВКИ

В результаті виконання курсової роботи було розроблено сховище даних російських військовослужбовців що беруть участь у війні проти України, що містить фактову таблицю та декілька вимірів. Реалізовано конвертацію даних з різних форматів, ETL процеси для завантаження даних, та графічний аналіз даних. Для реалізації поставленої задачі було використано PostgreSQL, мова програмування Python та застосунок Power BI.

На основі простого графічного аналізу було отримано наглядні графіки щодо загальної картини серед військових.

Отже, поставлені задачі були виконані, а також планується доповнення бази даних новою інформацією та підтримка проекту для подальшого введення в експлуатацію.

ПЕРЕЛІК ПОСИЛАНЬ

1. Керівництво користувача PD. Електронний ресурс режим доступу:
// <http://www.sybase.com/products/modelingdevelopment/powerdesigner>
2. . Гектор Гарсиа-Молина , Джеффри Д. Ульман , Дженнифер Уидом .
Системы баз данных.
Полный курс, 2004.1088с. ил. ISBN 5-8459-0384-X; 2004 г.
3. Барсегян А.А., Куприянов М.С. Степаненко В.В., Холод И.И. Методы и
модели анали за
данными: OLAP и Data Mining.- СПб.: БХВ-Петербург, 2004. 336 с.: ил. ISBN 5-
94157-522-X.
4. W. H. Inmon, Building the Data Warehouse, Fourth Edition: Wiley Publishing, Inc.
ISBN-13:
978-0-7645-9944-6, ISBN-10: 0-7645-9944-5
5. Jaspersoft OLAP Ultimate Guide. Навчальний посібник
6. Get started with Power BI Desktop - <https://docs.microsoft.com/en-us/power-bi/fundamentals/desktop-getting-started>

ДОДАТОК А ТЕКСТИ ПРОГРАМНОГО КОДУ

студента групи IT-03 II курсу

Чабана А.Є.

```

update "Льотчики РФ.xlsx - Лист1 (1)"
set "Имя" = trim("Имя", chr(160));

update "Льотчики РФ.xlsx - Лист1 (1)"
set "Фамилия" = trim("Фамилия", chr(160));

update "Льотчики РФ.xlsx - Лист1 (1)"
set "Отчество" = trim("Отчество", chr(160));

delete
from "Льотчики РФ.xlsx - Лист1 (1)"
where "Фамилия" is null;

select CONCAT("Фамилия", ' ', "Имя", ' ', "Отчество") as "fio",
       "Дата рождения",
       "Место рождения",
       "Национальность",
       "Семейное положение",
       "Адрес места жительства (населенны",
       "Воинское звание",
       "Образование",
       "Номер паспорта",
       "Дата выдачи паспорта"
from "Льотчики РФ.xlsx - Лист1 (1)";

insert into orc (fio, bdate, bplace, nationality, mstatus,
address, mil_rank, study_place, pass_num, pass_date)
select CONCAT("Фамилия", ' ', "Имя", ' ', "Отчество") as "fio",
       "Дата рождения",
       "Место рождения",
       "Национальность",
       "Семейное положение",
       "Адрес места жительства (населенны",
       "Воинское звание",
       "Образование",
       "Номер паспорта",
       "Дата выдачи паспорта"
from "Льотчики РФ.xlsx - Лист1 (1)";

update orc
set source = 'Льотчики РФ.xlsx'
where source is null;

-- #####

SELECT to_date(bdate, 'MM/DD/YYYY')
from "Войсковая часть 78018.xlsx - Лист1";

select "Фамилия, имя, отчество", bdate, "Воинское звание", "Личный
номер"

```

```

from "Войсковая часть 78018.xlsx - Лист1";

insert into orc (fio, bdate, mil_rank, lnum)
select "Фамилия, имя, отчество", bdate, "Воинское звание", "Личный
номер"
from "Войсковая часть 78018.xlsx - Лист1";

update orc
set source = 'Войсковая часть 78018.xlsx'
where source is null;

-- #####

SELECT to_date(bdate, 'DD.MM.YYYY')
from "Войсковая часть 77984.xlsx - Лист1";

update "Войсковая часть 77984.xlsx - Лист1"
set bdate = to_date(bdate, 'DD.MM.YYYY');

select "Фамилия, имя, отчество", bdate, "Воинское звание", "Личный
номер"
from "Войсковая часть 77984.xlsx - Лист1";

insert into orc (fio, bdate, mil_rank, lnum)
select "Фамилия, имя, отчество", bdate, "Воинское звание", "Личный
номер"
from "Войсковая часть 77984.xlsx - Лист1";

update orc
set source = 'Войсковая часть 77984.xlsx'
where source is null;

-- #####

SELECT to_date("ДатаРожд", 'MM.DD.YY')
from "rf200-1.xlsx - VK";

update "rf200-1.xlsx - VK"
set "ДатаРожд" = to_date("ДатаРожд", 'MM.DD.YY');

update "rf200-1.xlsx - VK"
set "ДатаРожд" = REPLACE(
    "ДатаРожд",
    '20',
    '19')
where "ДатаРожд" like '20%';

update "rf200-1.xlsx - VK"
set "Дата выдачи" = to_date("Дата выдачи", 'MM.DD.YY');

update "rf200-1.xlsx - VK"
set "Дата выдачи" = null

```

```

where "Дата выдачи" = '<null>';

update "rf200-1.xlsx - VK"
set "Наименование воинского звания" = null
where "Наименование воинского звания" like '1%'
or "Наименование воинского звания" like '2%'
or "Наименование воинского звания" like '3%'
or "Наименование воинского звания" like '4%'
or "Наименование воинского звания" like '5%'
or "Наименование воинского звания" like '6%'
or "Наименование воинского звания" like '7%'
or "Наименование воинского звания" like '8%'
or "Наименование воинского звания" like '9%'
or "Наименование воинского звания" like '0%';

update "rf200-1.xlsx - VK"
set "Подтип" = null
where "Подтип" not like 'Паспорт гражданина Российской Федерации';

update "rf200-1.xlsx - VK"
set "Серия" = null
where "Подтип" not like 'Паспорт гражданина Российской Федерации';

update "rf200-1.xlsx - VK"
set "Ид. номер" = null
where "Подтип" not like 'Паспорт гражданина Российской Федерации';

update "rf200-1.xlsx - VK"
set "Дата выдачи" = null
where "Подтип" not like 'Паспорт гражданина Российской Федерации';

update "rf200-1.xlsx - VK"
set "Кем выдан" = null
where "Подтип" not like 'Паспорт гражданина Российской Федерации';

select "ФИО",
       "ДатаРожд",
       "Ключ пола",
       "ВК",
       "Таб.№",
       "Наименование воинского звания",
       "Серия",
       "Ид. номер",
       "Дата выдачи",
       "Кем выдан"
from "rf200-1.xlsx - VK";

insert into orc (fio, bdate, gender, vk, tab_num, mil_rank,
pass_ser, pass_num, pass_date, pass_auth)
select "ФИО",
       "ДатаРожд",

```

```

        "Ключ пола",
        "ВК",
        "Таб.№",
        "Наименование воинского звания",
        "Серия",
        "Ид. номер",
        "Дата выдачи",
        "Кем выдан"
from "rf200-1.xlsx - VK";

update orc
set gender = 'm'
where orc.gender like 'н%';

update orc
set source = 'rf200-1.xlsx'
where source is null;

-- #####

update "Полонені РФ ВЧ 31135.xlsx - Аркуш1 (1)"
set "ДАТА НАРОДЖЕННЯ" = to_date("ДАТА НАРОДЖЕННЯ", 'DD.MM.YY');

insert into orc(fio, bplace, v_ch)
select "ПІВ", "МІСЦЕ НАРОДЖЕННЯ", "ВІЙСЬКОВА ЧАСТИНА"
from "Полонені РФ ВЧ 31135.xlsx - Аркуш1 (1)";

update orc
set source = 'Полонені РФ ВЧ 31135.xlsx'
where source is null;

UPDATE orc
SET bdate=subquery.bdate,
    bdate_cl=subquery.bdate,
    fio=subquery.fio,
    v_ch=subquery."Раздел персонала",
    tab_num=subquery.tabn,
    lnum=subquery.ln,
    mil_rank=subquery.mlrank,
    gender=subquery."Ключ пола",
    pass_ser=subquery."Серия",
    pass_num=subquery."Ид. номер",
    pass_auth=subquery."Кем выдан",
    pass_date=subquery."Дата выдачи",
    pass_date_cl=subquery."Дата выдачи",
    mil_rank_cl=subquery.mlrank
FROM (SELECT bdate,
             fio,
             "Раздел персонала",
             tabn,
             ln,
             mlrank,

```

```

        "Ключ пола",
        "Серия",
        "Ид. номер",
        "Кем выдан",
        "Дата выдачи"
FROM pdforcs) AS subquery
WHERE orc.fio = subquery.fio;

insert into orc (bdate, bdate_cl, fio, v_ch, tab_num, lnum,
mil_rank, gender, pass_ser, pass_num, pass_auth, pass_date,
pass_date_cl, age, mil_rank_cl)

select bdate,
bdate,
fio,
"Раздел персонала",
tabn,
ln,
mlrank,
"Ключ пола",
"Серия",
"Ид. номер",
"Кем выдан",
"Дата выдачи",
"Дата выдачи",
date_part('year', age(current_date, bdate::date))::int,
mlrank
from pdforcs;

select date_part('year', age(current_date, bdate::date))::int
from pdforcs;

```