Київський національний університет імені Тараса Шевченка

**Лабораторна робота №2**

**З навчального курсу «Розробка бізнес-аналітичних систем»**

**«** **Czech Financial Dataset»**

Виконав:

студент 4 курсу

факультету кібернетики

спеціальність «Комп’ютерні науки»

групи ТТП-42

Чебан Богдан Володимирович

**Київ 2025**

**Завдання (Task)**

**Мета:**  
**Розробка Інтерактивного Веб-Додатку для Всебічного Аналізу Бізнес-Даних**

Метою цього проекту є створення інтерактивного веб-додатку за допомогою Dash, який надасть користувачам можливість здійснювати всебічний аналіз бізнес-даних через різноманітні візуалізації та ключові показники ефективності (KPI). Додаток повинен бути всеохоплюючим та самодостатнім, забезпечуючи інтуїтивно зрозумілий інтерфейс, який дозволяє користувачам швидко оцінити стан бізнесу без необхідності додаткових пояснень або навчання.

**Основні Цілі:**

1. **Забезпечення Швидкого Доступу до Ключових Даних:**

Надання користувачам можливості швидко отримувати необхідну інформацію про різні аспекти бізнесу через зрозумілі графіки та KPI.

1. **Покращення Прийняття Рішень:**

Допомога керівникам та аналітикам у прийнятті обґрунтованих рішень на основі актуальних та точних даних.

1. **Підвищення Ефективності Роботи:**

Автоматизація процесу збору та аналізу даних, що дозволяє зекономити час та ресурси компанії.

1. **Гнучкість та Адаптивність:**

Забезпечення можливості користувачам фільтрувати та аналізувати дані згідно з їхніми потребами та інтересами.

**Основні Вимоги:**

1. **Розробка Dashboard:**

Створення Dashboard з різними типами графіків (мінімум 9), які відображають різні аспекти бізнес-даних, такі як фінансові транзакції, кредитні операції, демографічні дані клієнтів тощо.

1. **Ключові Показники Ефективності (KPI):**

Визначення та відображення більше ніж 4 KPI, які є критично важливими для оцінки бізнес-процесів, таких як загальна кількість транзакцій, середня сума транзакції, загальна сума кредитів, кількість унікальних клієнтів тощо.

1. **Таблиця зі Спарклайнами:**

Реалізація таблиці, яка включає спарклайни для візуалізації трендів даних, що дозволяє користувачам швидко оцінити зміни та тенденції у різних регіонах або сегментах бізнесу.

1. **Інтерактивність:**

Забезпечення можливості фільтрації даних за регіоном та віком клієнтів, що динамічно оновлює всі графіки та KPI. Інші можливі фільтри можуть включати тип транзакції, дистрикт тощо.

1. **Візуальна Привабливість:**

Застосування професійних стилів та шаблонів для забезпечення чистого, зрозумілого та естетично привабливого вигляду Dashboard. Використання адаптивного дизайну для оптимального відображення на різних пристроях.

1. **Оптимізація Продуктивності:**

Забезпечення швидкого завантаження та оновлення графіків та таблиць, навіть при роботі з великими обсягами даних.

1. **Безпека та Конфіденційність:**

Захист даних та забезпечення конфіденційності інформації про клієнтів та фінансові операції.

1. **Масштабованість:**

Розробка додатку з урахуванням можливості подальшого розширення функціональності та інтеграції з іншими системами або джерелами даних.

**Підходи до Вирішення Завдання:**

1. **Використання Dash як Фреймворку:**

Dash є потужним інструментом для створення інтерактивних веб-додатків на основі Python. Він дозволяє легко інтегрувати різноманітні візуалізації та забезпечувати високу ступінь інтерактивності.

1. **Агрегація та Обробка Даних:**

Використання pandas для обробки та агрегування даних, що дозволяє створювати необхідні набори даних для графіків та KPI.

1. **Візуалізація Даних:**

Використання plotly.express та plotly.graph\_objects для створення різноманітних графіків, які є інтерактивними та легко адаптуються до потреб користувача.

1. **Інтеграція Прогрес-Бару:**

Використання tqdm для відображення прогресу при генерації спарклайнів, що покращує користувацький досвід та інформує про стан процесу.

1. **Оптимізація Коду:**

Забезпечення оптимальної продуктивності додатку шляхом ефективного використання ресурсів та мінімізації часу обробки даних.

**Короткий Опис Домену (Domain Short Description)**

**Бізнес-Аналіз у Банківській Сфері**

**Огляд Домену:**

Банківська сфера є однією з найбільш динамічних та конкурентних галузей, де ефективний аналіз даних відіграє ключову роль у прийнятті стратегічних рішень. Фінансові установи використовують бізнес-аналітику для відстеження операцій, оцінки ризиків, покращення обслуговування клієнтів та оптимізації внутрішніх процесів. В умовах постійних змін на ринку та зростаючих вимог до прозорості та ефективності, можливість швидко аналізувати великі обсяги даних стає невід’ємною частиною успішної стратегії банку.

**Основні Аспекти Бізнес-Аналізу в Банківській Сфері:**

1. **Фінансові Транзакції:**

Аналіз транзакцій дозволяє відстежувати обсяги операцій, виявляти тенденції, визначати найбільш популярні продукти та послуги, а також виявляти аномалії, що можуть свідчити про шахрайство або інші проблеми.

1. **Кредитні Операції:**

Аналіз кредитних даних допомагає оцінювати ризики, визначати кредитоспроможність клієнтів, оптимізувати портфель кредитів та розробляти нові кредитні продукти.

1. **Демографічні Дані Клієнтів:**

Вивчення демографічних характеристик клієнтів (вік, стать, місцезнаходження, доходи тощо) дозволяє банкам краще розуміти свою клієнтську базу, персоналізувати пропозиції та покращувати клієнтський досвід.

1. **Операційна Ефективність:**

Аналіз внутрішніх процесів допомагає виявляти неефективності, оптимізувати ресурси та підвищувати продуктивність співробітників.

1. **Відстеження Показників Ефективності (KPI):**

Визначення та моніторинг KPI дозволяє оцінювати успішність банківських операцій, встановлювати цілі та відстежувати прогрес у досягненні стратегічних цілей.

**Виклики Бізнес-Аналізу у Банківській Сфері:**

1. **Обсяг та Складність Даних:**

Банки оперують великими обсягами даних, що включають як структуровані, так і неструктуровані дані. Обробка та аналіз таких даних вимагає потужних інструментів та ефективних методів обробки.

1. **Конфіденційність та Безпека:**

Банківські дані є високочутливими та конфіденційними. Забезпечення їх безпеки та відповідність нормативним вимогам є критично важливим аспектом бізнес-аналітики.

1. **Інтеграція Даних:**

Об'єднання даних з різних джерел (CRM-системи, фінансові системи, соціальні мережі тощо) може бути складним завданням через різні формати даних та стандарти.

1. **Точність та Якість Даних:**

Забезпечення високої якості даних є ключовим для точності аналізу. Некоректні або неповні дані можуть призвести до хибних висновків та рішень.

**Використання Інтерактивних Графіків та KPI:**

Інтерактивні графіки та KPI дозволяють користувачам отримувати глибоке розуміння даних у режимі реального часу. Вони забезпечують візуальне представлення складних даних, що спрощує процес прийняття рішень. Наприклад:

**Bar Charts:** Відображення суми транзакцій за регіонами допомагає визначити найбільш активні ринки.

**Line Charts:** Аналіз транзакцій по роках дозволяє відстежувати зростання або спад операцій.

**Pie Charts:** Розподіл типів транзакцій дає уявлення про популярні продукти та послуги.

**Scatter Charts:** Виявлення взаємозв'язку між сумою транзакцій та кредитами допомагає оцінювати ризики.

**Переваги Інтерактивних Візуалізацій:**

1. **Зручність Використання:**

Інтерактивні графіки дозволяють користувачам взаємодіяти з даними, наприклад, через наведення курсора, масштабування або фільтрацію, що підвищує зручність використання додатку.

1. **Швидкий Доступ до Інформації:**

Користувачі можуть швидко отримувати необхідну інформацію без потреби переглядати великі таблиці даних.

1. **Гнучкість Аналізу:**
   * Інтерактивні інструменти дозволяють проводити аналіз з різних кутів, що допомагає виявляти нові тенденції та інсайти.
2. **Візуальна Привабливість:**
   * Яскраві та добре оформлені графіки роблять процес аналізу даних більш привабливим та мотивуючим для користувачів.

## **Original Dataset Link + Short Description (Оригінальне посилання на набір даних + короткий опис)**

### ****English:****

* **Dataset:** 1999 Czech Bank Financial Dataset
* **Link:** <https://www.kaggle.com/datasets/marceloventura/the-berka-dataset/data>
* **Description:**  
  The **1999 Czech Bank Financial Dataset**, also known as **The Berka Dataset**, is a comprehensive collection of financial information from a Czech bank. Compiled by Petr Berka and Marta Sochorova, the dataset encompasses data related to over 5,300 bank clients and approximately 1,000,000 transactions conducted in the year 1999. Additionally, the dataset includes information on nearly 700 loans and about 900 credit cards issued to the clients.

**Key Components:**

**Transaction Data (trans.csv):**

* **Fields:**
* trans\_id: Transaction Identifier (Primary Key)
* date: Date of Transaction (Format: YYMMDD)
* account\_id: Account Identifier (Foreign Key to dim\_account)
* amount: Transaction Amount
* balance: Account Balance After Transaction
* client\_id: Client Identifier (Foreign Key to dim\_client)
* district\_id: District Identifier (Foreign Key to dim\_district)
* type\_of\_trans: Type of Transaction
* operation: Operation Details
* Additional fields related to loans and credit cards.

**Client Data (client.csv):**

* **Fields:**
* client\_id: Client Identifier (Primary Key)
* birth\_number: Client's Birth Number
* district\_id: District Identifier (Foreign Key to dim\_district)
* gender: Client's Gender
* birth\_date: Client's Birth Date
* Additional demographic and personal attributes.

**Account Data (account.csv):**

* **Fields:**
* account\_id: Account Identifier (Primary Key)
* frequency: Transaction Frequency
* date\_created: Account Creation Date
* Additional account-related attributes.

**District Data (district.csv):**

* **Fields:**
* district\_id: District Identifier (Primary Key)
* region: Geographic Region
* district\_name: Name of the District
* Additional geographic and administrative attributes.

**Order Data (order.csv):**

* **Fields:**
* order\_id: Order Identifier (Primary Key)
* client\_id: Client Identifier (Foreign Key to dim\_client)
* order\_date\_id: Order Date Identifier (Foreign Key to dim\_date)
* order\_amount: Amount of the Order
* Additional order-related measures.

**Dispensation Data (disp.csv):**

* **Fields:**
* disp\_id: Dispensation Identifier (Primary Key)
* client\_id: Client Identifier (Foreign Key to dim\_client)
* disp\_type: Type of Dispensation
* disp\_date\_id: Dispensation Date Identifier (Foreign Key to dim\_date)
* Additional dispensation-related measures.

**Loan Data (loan.csv):**

* **Fields:**
* loan\_id: Loan Identifier (Primary Key)
* client\_id: Client Identifier (Foreign Key to dim\_client)
* loan\_date\_id: Loan Date Identifier (Foreign Key to dim\_date)
* loan\_amount: Amount of the Loan
* loan\_type: Type of Loan
* Additional loan-related measures.

**Credit Card Data (card.csv):**

* **Fields:**
* card\_id: Credit Card Identifier (Primary Key)
* client\_id: Client Identifier (Foreign Key to dim\_client)
* card\_type: Type of Credit Card
* card\_issue\_date: Issue Date of the Credit Card
* card\_limit: Credit Limit of the Card
* Additional credit card-related measures.

**Additional Information:**

* + **Data Volume:** The dataset includes over 1,000,000 transaction records, providing a robust foundation for extensive financial analysis.
  + **Dimensions and Measures:** The data is structured to support a multi-dimensional analysis with four primary Dimension Tables (dim\_date, dim\_client, dim\_district, dim\_account) and multiple Fact Tables (fact\_trans, fact\_order, fact\_disp, fact\_loan, fact\_card) capturing various financial measures.
  + **Hierarchies:** The dim\_date table incorporates hierarchies such as year, quarter, month, and day, facilitating time-based analyses. Similarly, geographic hierarchies are present in the dim\_district table.
  + **Usage:** This dataset is ideal for constructing an OLAP warehouse using a star schema, enabling efficient querying, reporting, and visualization of financial data for strategic decision-making.

### ****Українська:****

* **Набір Даних:** Фінансовий набір даних чеського банку за 1999 рік
* **Посилання:** <https://www.kaggle.com/datasets/marceloventura/the-berka-dataset/data>
* **Опис:**  
  **Фінансовий набір даних чеського банку за 1999 рік**, також відомий як **The Berka Dataset**, є всебічним збором фінансової інформації від чеського банку. Складаний Петром Беркою та Мартією Сохровою, набір даних охоплює інформацію про понад 5,300 клієнтів банку та близько 1,000,000 транзакцій, проведених у 1999 році. Крім того, набір даних включає інформацію про майже 700 кредитів та близько 900 кредитних карток, виданих клієнтам.

**Ключові компоненти:**

**Транзакційні Дані (trans.csv):**

* **Поля:**
* trans\_id: Ідентифікатор транзакції (Первинний ключ)
* date: Дата транзакції (Формат: YYMMDD)
* account\_id: Ідентифікатор рахунку (Зовнішній ключ до dim\_account)
* amount: Сума транзакції
* balance: Баланс рахунку після транзакції
* client\_id: Ідентифікатор клієнта (Зовнішній ключ до dim\_client)
* district\_id: Ідентифікатор району (Зовнішній ключ до dim\_district)
* type\_of\_trans: Тип транзакції
* operation: Деталі операції
* Додаткові поля, пов'язані з кредитами та кредитними картками.

**Дані про Клієнтів (client.csv):**

* **Поля:**
* client\_id: Ідентифікатор клієнта (Первинний ключ)
* birth\_number: Номер народження клієнта
* district\_id: Ідентифікатор району (Зовнішній ключ до dim\_district)
* gender: Стать клієнта
* birth\_date: Дата народження клієнта
* Додаткові демографічні та персональні атрибути.

**Дані про Рахунки (account.csv):**

* **Поля:**
* account\_id: Ідентифікатор рахунку (Первинний ключ)
* frequency: Частота транзакцій
* date\_created: Дата створення рахунку
* Додаткові атрибути, пов'язані з рахунками.

**Дані про Райони (district.csv):**

* **Поля:**
* district\_id: Ідентифікатор району (Первинний ключ)
* region: Географічний регіон
* district\_name: Назва району
* Додаткові географічні та адміністративні атрибути.

**Дані про Замовлення (order.csv):**

* **Поля:**
* order\_id: Ідентифікатор замовлення (Первинний ключ)
* client\_id: Ідентифікатор клієнта (Зовнішній ключ до dim\_client)
* order\_date\_id: Ідентифікатор дати замовлення (Зовнішній ключ до dim\_date)
* order\_amount: Сума замовлення
* Додаткові міри, пов'язані з замовленнями.

**Дані про Диспетчеризацію (disp.csv):**

* **Поля:**
* disp\_id: Ідентифікатор диспетчеризації (Первинний ключ)
* client\_id: Ідентифікатор клієнта (Зовнішній ключ до dim\_client)
* disp\_type: Тип диспетчеризації
* disp\_date\_id: Ідентифікатор дати диспетчеризації (Зовнішній ключ до dim\_date)
* Додаткові міри, пов'язані з диспетчеризацією.

**Дані про Кредити (loan.csv):**

* **Поля:**
* loan\_id: Ідентифікатор кредиту (Первинний ключ)
* client\_id: Ідентифікатор клієнта (Зовнішній ключ до dim\_client)
* loan\_date\_id: Ідентифікатор дати кредиту (Зовнішній ключ до dim\_date)
* loan\_amount: Сума кредиту
* loan\_type: Тип кредиту
* Додаткові міри, пов'язані з кредитами.

**Дані про Кредитні Картки (card.csv):**

* **Поля:**
* card\_id: Ідентифікатор кредитної картки (Первинний ключ)
* client\_id: Ідентифікатор клієнта (Зовнішній ключ до dim\_client)
* card\_type: Тип кредитної картки
* card\_issue\_date: Дата випуску кредитної картки
* card\_limit: Ліміт кредитної картки
* Додаткові міри, пов'язані з кредитними картками.

**Додаткова інформація:**

* + **Обсяг Даних:** Набір даних включає понад 1,000,000 записів транзакцій, що забезпечує міцну основу для детального фінансового аналізу.
  + **Виміри та Міри:** Дані структуровані для підтримки багатовимірного аналізу з чотирма основними Dimension Tables (dim\_date, dim\_client, dim\_district, dim\_account) та кількома Fact Tables (fact\_trans, fact\_order, fact\_disp, fact\_loan, fact\_card), що фіксують різні фінансові міри.
  + **Ієрархії:** Таблиця dim\_date включає ієрархії, такі як рік, квартал, місяць та день, що сприяє аналізу за часовими ознаками. Аналогічно, географічні ієрархії присутні в таблиці dim\_district.

### ****Методологія (Methodology)****

**Етапи Реалізації:**

1. **Ініціалізація та Підготовка Даних:**

Генерація штучних даних для трьох основних DataFrame (fact\_trans, fact\_loan, dim\_client).

Агрегування даних для створення різних наборів даних, необхідних для графіків та KPI.

1. **Обчислення KPI:**

Визначення ключових показників ефективності, таких як загальна кількість транзакцій, середня сума транзакції, загальна сума кредитів, кількість унікальних клієнтів тощо.

1. **Підготовка Даних для Графіків:**

Створення агрегованих DataFrame для різних типів графіків, включаючи бар-чарти, лінійні графіки, кругові діаграми, гістограми тощо.

1. **Створення Таблиці зі Спарклайнами:**

Розробка функції для генерації SVG-зображень спарклайнів.

Генерація спарклайнів для топ-10 регіонів за транзакціями та вставлення їх у таблицю.

1. **Розробка Dash-Додатку:**

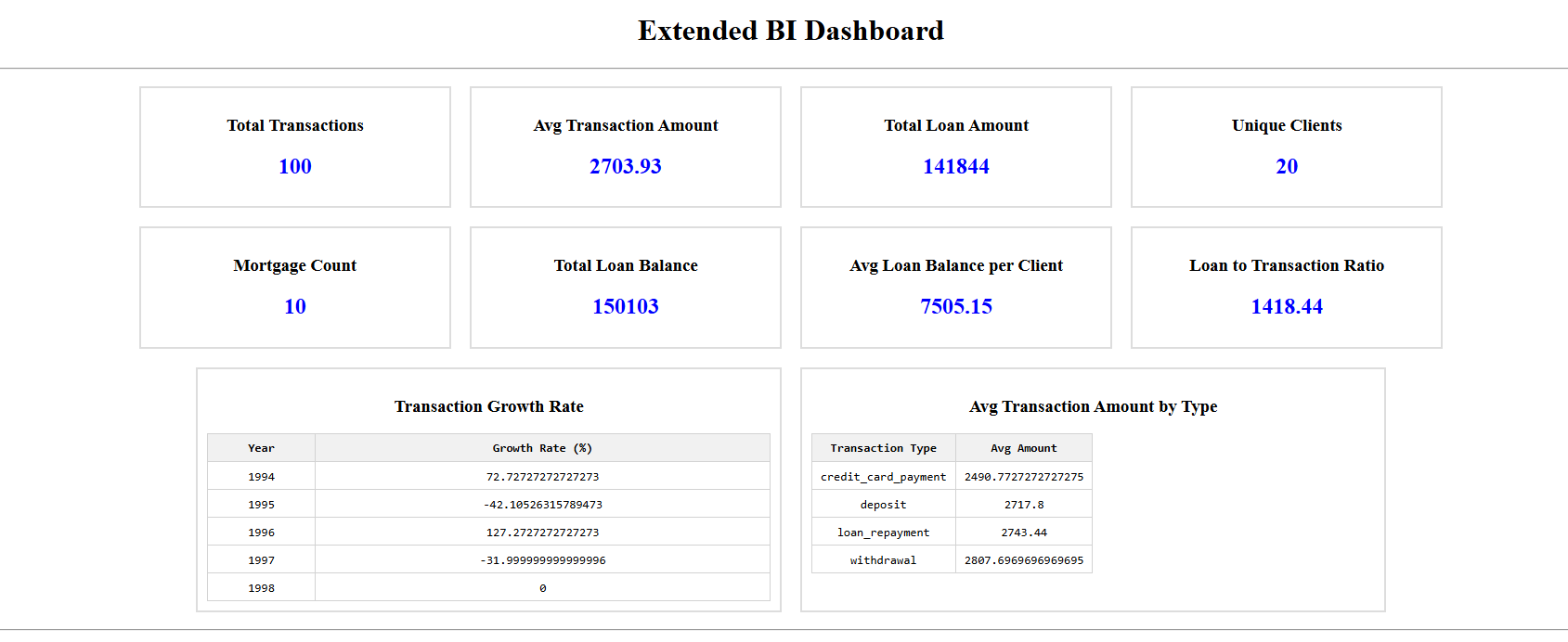
Ініціалізація Dash-додатку та визначення його макету (layout).

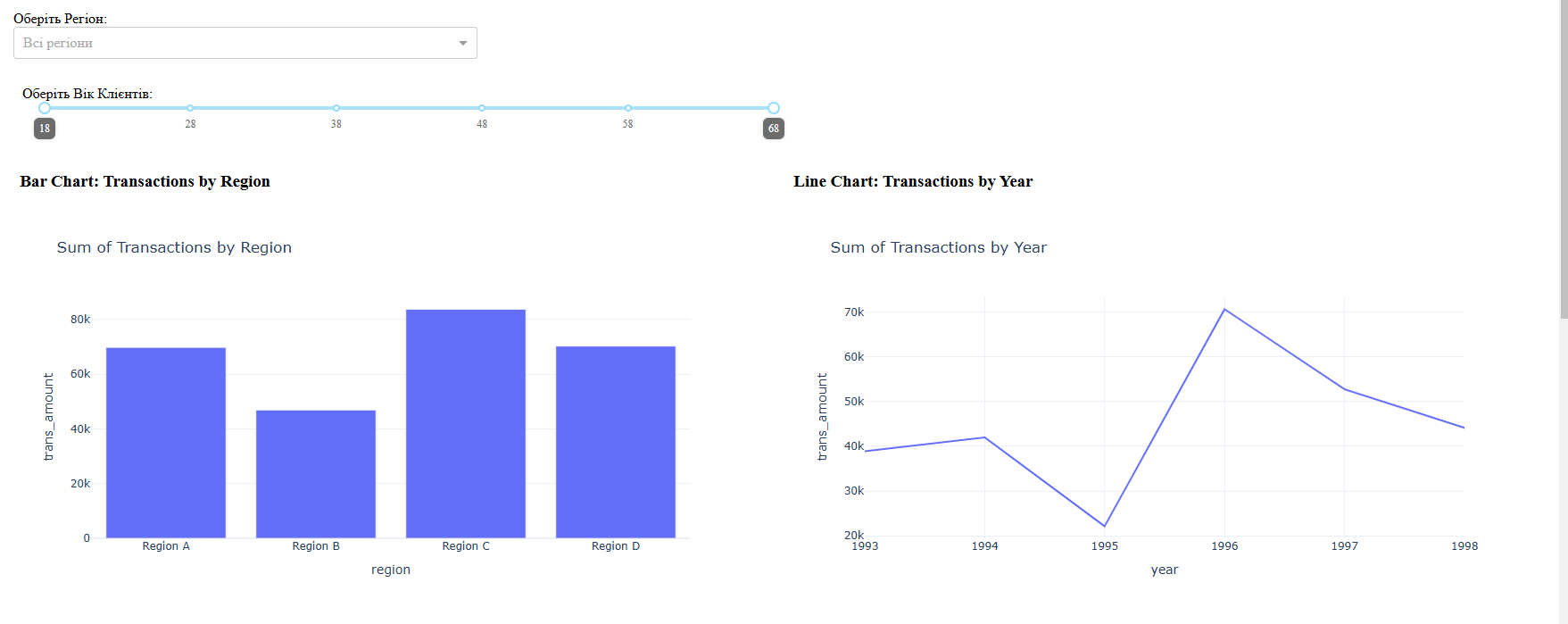
Розміщення KPI-карт, графіків, фільтрів та таблиці зі спарклайнами у відповідних секціях Dashboard.

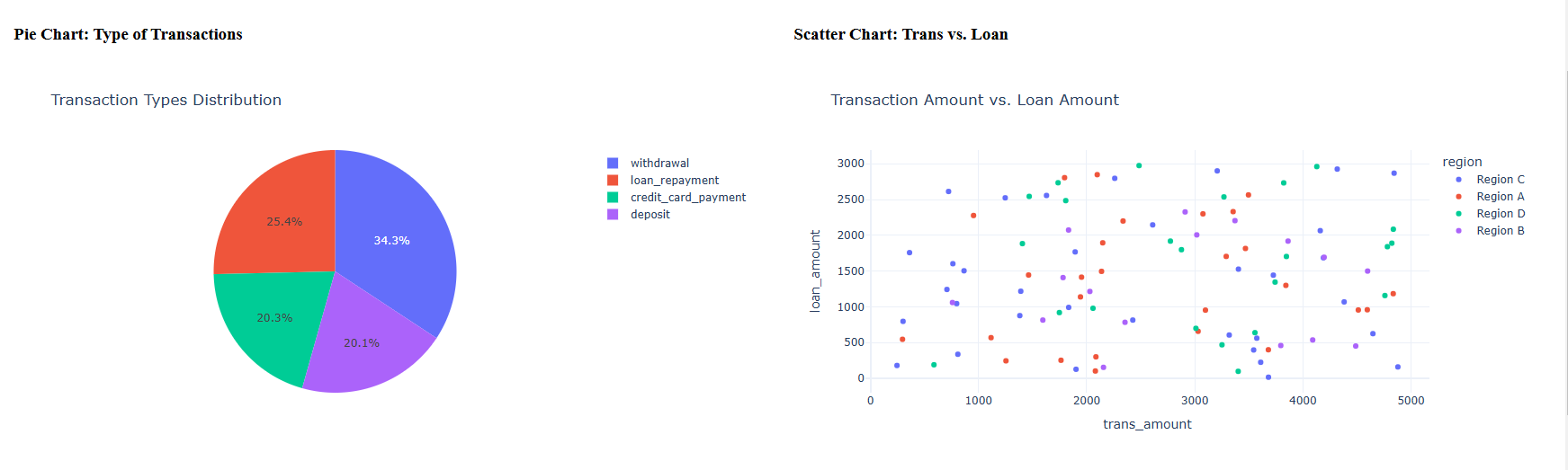
1. **Визначення Callback-Функцій:**

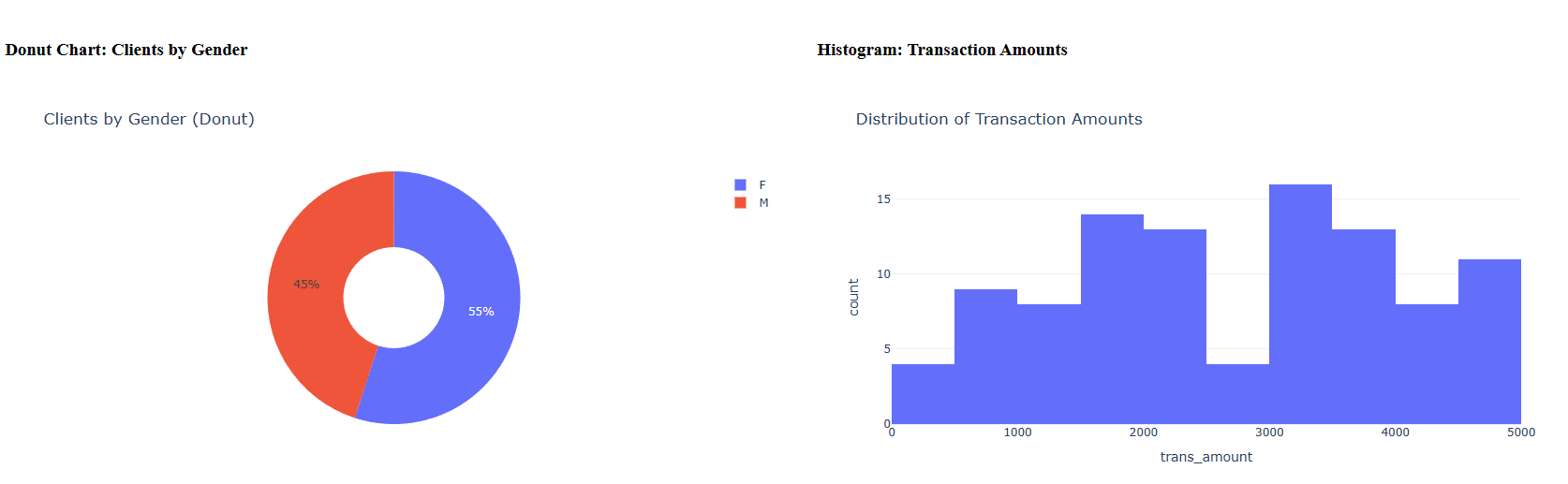
Реалізація інтерактивності через callback-функції, які оновлюють графіки та KPI відповідно до вибраних фільтрів.

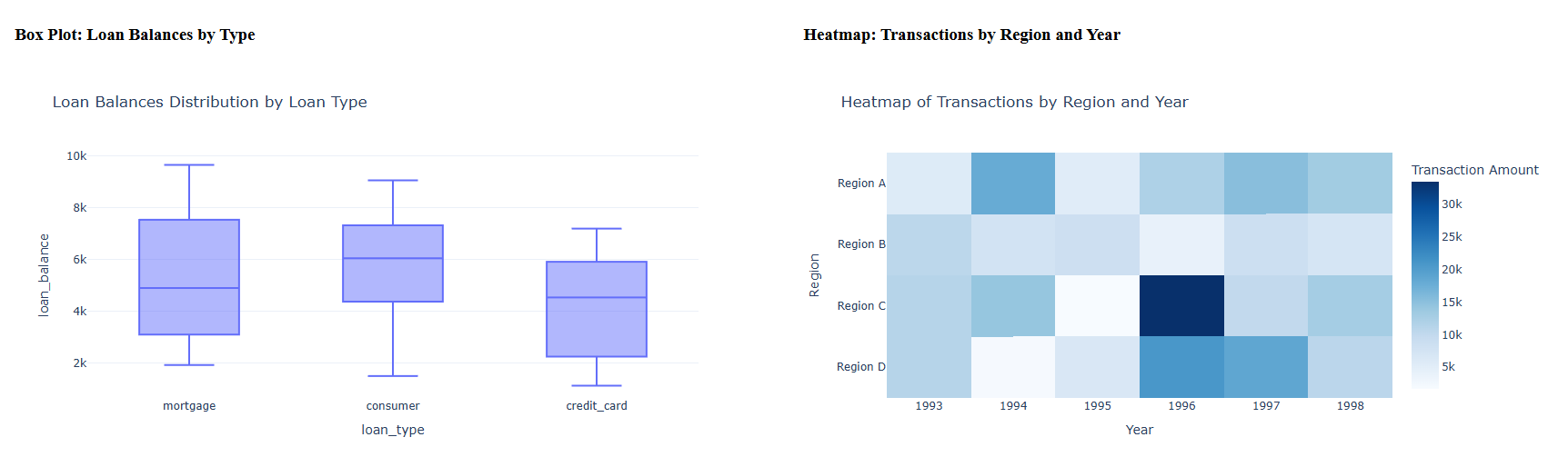
**Результати та Висновки (Results and Conclusions)**

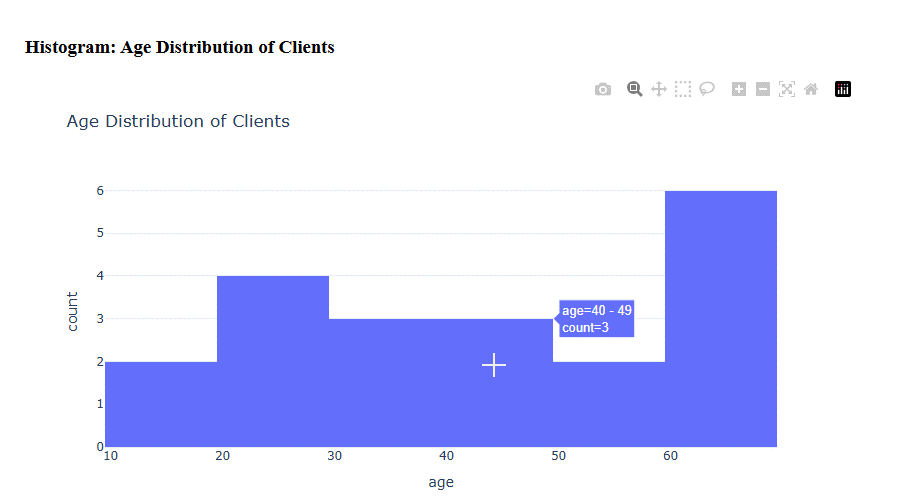


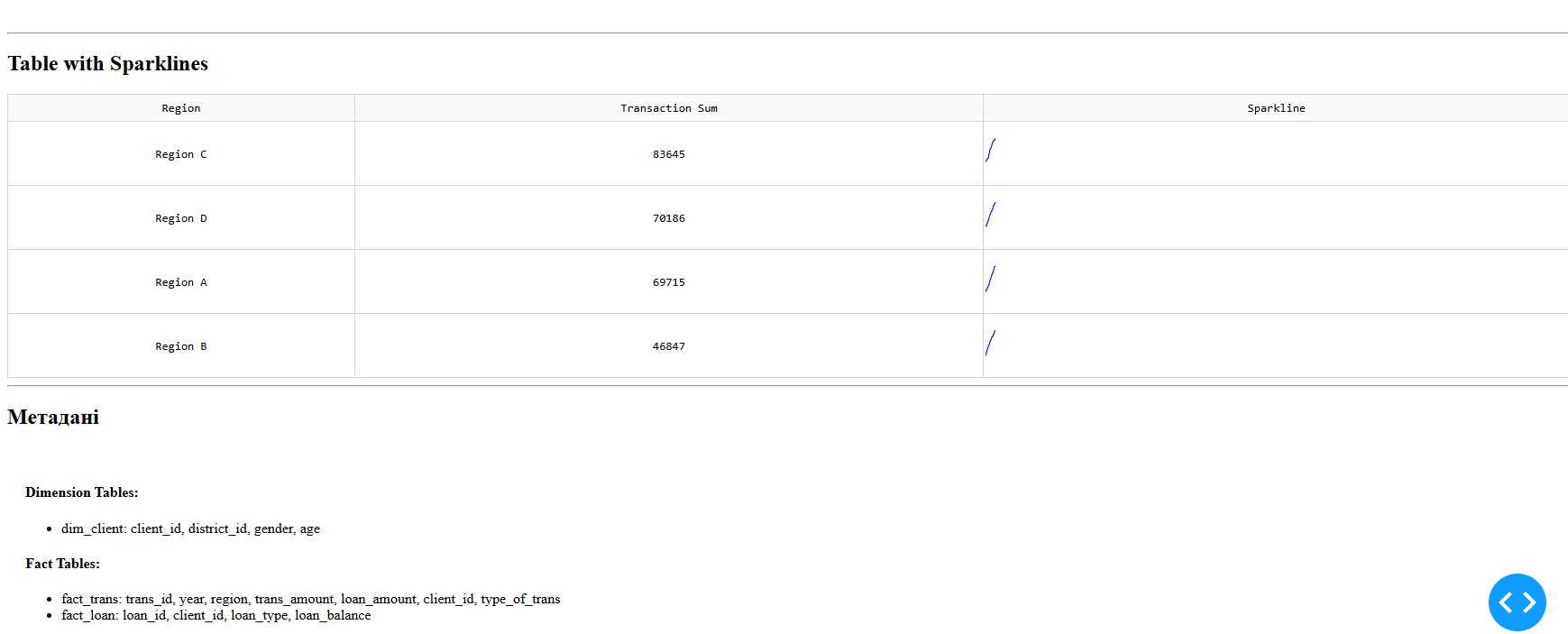












**Досягнуті Результати:**

* **Всебічний Аналіз:** Створений Dash-додаток забезпечує всебічний аналіз бізнес-даних через різноманітні графіки та KPI, що дозволяє користувачам швидко оцінювати стан бізнесу.
* **Інтерактивність:** Фільтри за регіоном та віком клієнтів забезпечують гнучкість аналізу та дозволяють користувачам зосередитися на релевантних даних.
* **Візуальна Привабливість:** Використання чистого та професійного шаблону plotly\_white разом із простими CSS-стилями робить Dashboard зрозумілим та естетично привабливим.
* **Оптимізація:** Генерація спарклайнів із використанням прогрес-бару дозволяє відстежувати процес створення графіків, що особливо корисно при роботі з великими обсягами даних.

### ****Висновок****

Створений Dash-додаток успішно виконує свою роль як інструмент для всебічного бізнес-аналізу у банківській сфері. Він об'єднує різноманітні графіки, KPI та таблиці зі спарклайнами у зручний та інтуїтивно зрозумілий інтерфейс, що дозволяє користувачам швидко отримувати необхідну інформацію та проводити гнучкий аналіз даних. Інтерактивність та висока візуальна привабливість забезпечують позитивний користувацький досвід, а масштабованість та гнучкість архітектури дозволяють додатку адаптуватися до змінних потреб бізнесу.