

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

Інститут прикладної математики та фундаментальних наук

Кафедра прикладної математики

Звіт про виконання курсової роботи з курсу
«Надвеликі бази даних»

Виконав:
Студент групи ПМ-43
Музика Денис
Перевірив:
Любінський Б.Б.

Львів-2024

1. Завдання.

Мета роботи полягає у закріпленні та поглибленні теоретичних знань і навичок, отриманих під час вивчення лекційного курсу та виконання лабораторних практикумів, а також у розвитку умінь самостійної творчої роботи шляхом аналізу власної великої бази даних. Основними завданнями роботи є створення звітів, побудованих на основі OLAP-кубів, за допомогою сучасних інструментів: SQL Server Management Studio (SSMS) та Visual Studio.

Завдання курсової роботи:

1. Вивчити основи роботи з OLAP-кубами в SQL Server Management Studio (SSMS) та Visual Studio.
2. Провести аналіз предметної області, на основі якої створюється база даних і OLAP-куби.
3. Створити багатовимірну модель даних із використанням SQL Server Analysis Services (SSAS).
4. Розробити звіти на основі створених кубів, які дозволяють проводити аналітичний аналіз даних.
5. Забезпечити візуалізацію аналітичних даних через інтеграцію зі сторонніми інструментами (Excel, Power BI тощо).

Зміст

1. Завдання.
2. Вступ.
3. Основна частина:
 - 3.1. Створення бази даних.
 - 3.2. Проектування OLAP-куба.
 - 3.3. Генерація звітів.
4. Висновки.
5. Джерела.

2. Вступ

OLAP (On-Line Analytical Processing) є невід'ємною складовою сучасних систем бізнес-аналітики, забезпечуючи можливість ефективної роботи з великими обсягами даних. Ця технологія дозволяє користувачам швидко отримувати відповіді на складні запити, які стосуються багатовимірного аналізу даних. OLAP забезпечує аналіз даних у розрізі різних вимірів, таких як час, регіон, категорія продуктів або інші бізнес-параметри.

Основна перевага OLAP полягає у швидкій обробці даних, навіть якщо вони містяться у великих базах даних. Наприклад, організації можуть використовувати OLAP для оцінки ефективності продажів у різних регіонах, визначення найпопулярніших продуктів чи аналізу поведінки споживачів у певний період.

Багатовимірний аналіз даних в OLAP реалізується через структуру, яка називається OLAP-кубом. OLAP-куби представляють дані у багатовимірному вигляді, де кожен вимір відповідає певному бізнес-параметру, а агрегації забезпечують підсумкові значення, наприклад, загальний обсяг продажів або середню ціну товару.

Значення OLAP у сучасних системах бізнес-аналітики

Системи бізнес-аналітики дозволяють підприємствам приймати зважені рішення на основі аналізу великих обсягів даних. OLAP є основою для таких систем, оскільки забезпечує:

1. **Швидкий доступ до даних:** Завдяки попередньо обчисленим агрегатам, OLAP дозволяє отримувати результати запитів у лічені секунди.
2. **Підтримку складного аналізу:** OLAP дозволяє виконувати багатовимірні запити, які важко реалізувати за допомогою традиційних реляційних баз даних.
3. **Гнучкість:** Користувачі можуть виконувати аналіз за різними параметрами, змінювати рівень деталізації даних, переходячи від загального огляду до детальних даних.

OLAP особливо корисний для великих організацій, які працюють із великими масивами даних, наприклад, у фінансах, торгівлі чи маркетингу. Він дозволяє ідентифікувати тренди, виявляти закономірності та робити прогнози на основі історичних даних.

SQL Server Analysis Services (SSAS)

SQL Server Analysis Services (SSAS) є компонентом SQL Server, який надає повний набір інструментів для створення та управління багатовимірними моделями даних. Основні функції SSAS включають:

1. **Створення OLAP-кубів:** SSAS дозволяє проектувати багатовимірні моделі, які можуть включати десятки вимірів і сотні мір для аналізу.
2. **Управління агрегаціями:** Для підвищення продуктивності SSAS автоматично створює агрегати, які забезпечують швидкий доступ до узагальнених даних.

3. **Підтримка багатовимірних запитів (MDX):** Мова MDX (MultiDimensional eXpressions) використовується для виконання складних запитів до OLAP-кубів.
4. **Інтеграція:** SSAS інтегрується з іншими продуктами Microsoft, такими як Excel та Power BI, що дозволяє легко створювати візуалізації даних.

SSAS також має інструменти для управління ролями користувачів і забезпечення безпеки даних, що особливо важливо для підприємств із конфіденційними даними. Завдяки своїй гнучкості та потужності, SSAS є вибором багатьох організацій для побудови систем бізнес-аналітики.

Visual Studio як інструмент для роботи з OLAP

Visual Studio є потужним середовищем розробки, яке використовується для створення програмних продуктів, включаючи BI-проекти. Воно забезпечує:

1. **Проектування OLAP-кубів:** У Visual Studio можна легко налаштувати джерела даних, створити виміри та міри, визначити схеми агрегації.
2. **Налагодження моделей:** Вбудовані інструменти для тестування дозволяють перевіряти коректність роботи OLAP-кубів ще до їх розгортання.
3. **Інтеграція з SSAS:** Visual Studio дозволяє розгорнути OLAP-куби на сервері SSAS, що забезпечує їх готовність до використання в реальному середовищі.
4. **Гнучкість:** Інструмент підтримує інтеграцію зі сторонніми додатками, такими як Excel, для створення звітів та візуалізації даних.

Завдяки Visual Studio процес створення OLAP-кубів стає інтуїтивним і структурованим. Розробники отримують доступ до візуальних інструментів для управління багатовимірними моделями, що значно спрощує процес їх розробки.

Значення поєднання SSAS та Visual Studio

Поєднання SSAS та Visual Studio забезпечує потужний і гнучкий інструментарій для роботи з OLAP. Використовуючи ці інструменти, можна:

- Автоматизувати процес створення моделей даних.
- Забезпечити високу продуктивність системи через оптимізацію агрегацій.
- Створювати звіти, які відповідають потребам бізнесу.
- Здійснювати інтеграцію з іншими продуктами Microsoft для покращення бізнес-аналітики.

Таким чином, OLAP, SSAS і Visual Studio забезпечують комплексне рішення для створення ефективних систем аналітики, які допомагають підприємствам отримувати цінну інформацію з великих обсягів даних. У рамках цієї роботи ці інструменти використовуються для побудови системи звітності, яка відповідає сучасним вимогам бізнес-аналітики.

3. Основна частина

Ось оновлений детальний опис процесу з урахуванням того, що база даних називається **Bank**, а максимальний кредитний рейтинг дорівнює 100:

3.1. Створення бази даних

Створення бази даних є першим етапом роботи, що закладає основу для аналізу кредитного рейтингу клієнтів. У цій роботі використовується база даних із назвою **Bank**, у якій кредитний рейтинг клієнтів варіюється в межах від 0 до 100.

1. Планування структури бази даних:

База даних складається з двох основних таблиць:

- **Клієнти (Clients):** містить інформацію про кожного клієнта банку, включаючи:
 - ClientID — унікальний ідентифікатор клієнта.
 - FirstName і LastName — ім'я та прізвище клієнта.
 - CreditRating — кредитний рейтинг клієнта (числове значення в межах від 0 до 100).
- **Транзакції (Transactions):** записує фінансові операції клієнтів, включаючи:
 - TransactionID — унікальний ідентифікатор транзакції.
 - ClientID — ідентифікатор клієнта, якому належить транзакція.
 - TransactionDate — дата транзакції.
 - Amount — сума транзакції.
 - TransactionType — тип транзакції (наприклад, "Deposit" або "Withdrawal").

Зв'язок між таблицями реалізовано через поле ClientID.

2. Створення бази даних у SQL Server Management Studio (SSMS):

Написав SQL-запити для створення бази даних та основних таблиць

Recordbook	First_Name	Last_Name	Date_of_Birth	Enter_Date
1000435	Krysten	Keesler	1992-02-20	2010-11-04
1000436	Lovetta	Nicka	1993-01-23	2013-05-24
1000437	Azucena	Quaye	1992-02-05	2009-01-24
1000438	Chanelle	Manjarrez	1993-11-26	2011-04-08
1000439	Drucilla	Yoast	1993-02-23	2011-09-02
1000440	Francisco	Sudak	1992-03-27	2014-09-22
1000441	Jasper	Clemenza	1993-09-19	2012-11-14
1000442	Piedad	Westry	1991-07-03	2011-03-27
1000443	Lisabeth	Squyres	1991-05-10	2011-04-29
1000444	Milford	Shawhan	1993-10-30	2009-06-07
1000445	Elease	Hutchins	1992-09-10	2013-04-06
1000446	Crysta	Ingegneri	1992-12-14	2014-04-09
1000447	Tracey	Maddix	1991-06-27	2013-08-28
1000448	Vincenzo	Susanin	1993-05-09	2013-03-01
1000449	Lexie	Quillen	1993-08-14	2009-02-23
1000450	Sigrid	Alsup	1991-06-27	2010-11-08
1000451	Nora	Honn	1991-04-08	2013-08-15
1000452	Lacy	Perciful	1993-07-15	2009-01-09
1000453	Dorcas	Cogen	1991-04-08	2009-01-17
1000454	Thaddeus	Menck	1991-10-31	2014-10-02
1000455	Rosenda	Bahn	1991-10-27	2012-01-14
1000456	Ricarda	Szot	1993-12-02	2009-05-27
1000457	Mamie	Ahne	1992-03-06	2013-09-02
1000458	Lorna	Deeters	1993-06-23	2011-08-28
1000459	Gregory	Squair	1992-01-03	2013-12-06
1000460	Tonia	Bachtell	1991-06-10	2010-07-12
1000461	Anna	Holstein	1993-04-18	2009-09-28
1000462	Leeann	Kingma	1992-02-06	2010-08-18
1000463	Melania	Nibbs	1992-11-06	2012-07-17
1000464	Della	Sterr	1992-06-21	2009-09-25
1000465	Jami	Phillipson	1993-11-29	2010-06-28
1000466	Elidia	Miramont...	1993-10-28	2010-08-21
1000467	Jed	Sigona	1991-08-25	2012-08-24
1000468	Lucina	Bellido	1992-07-28	2014-12-02
1000469	Hallie	Doeberling	1991-04-20	2014-10-17
1000470	Kerry	Kreinbring	1992-05-13	2011-07-27
1000471	Dominic	Mehlman	1993-10-27	2010-09-23

3.2. Проектування OLAP-куба

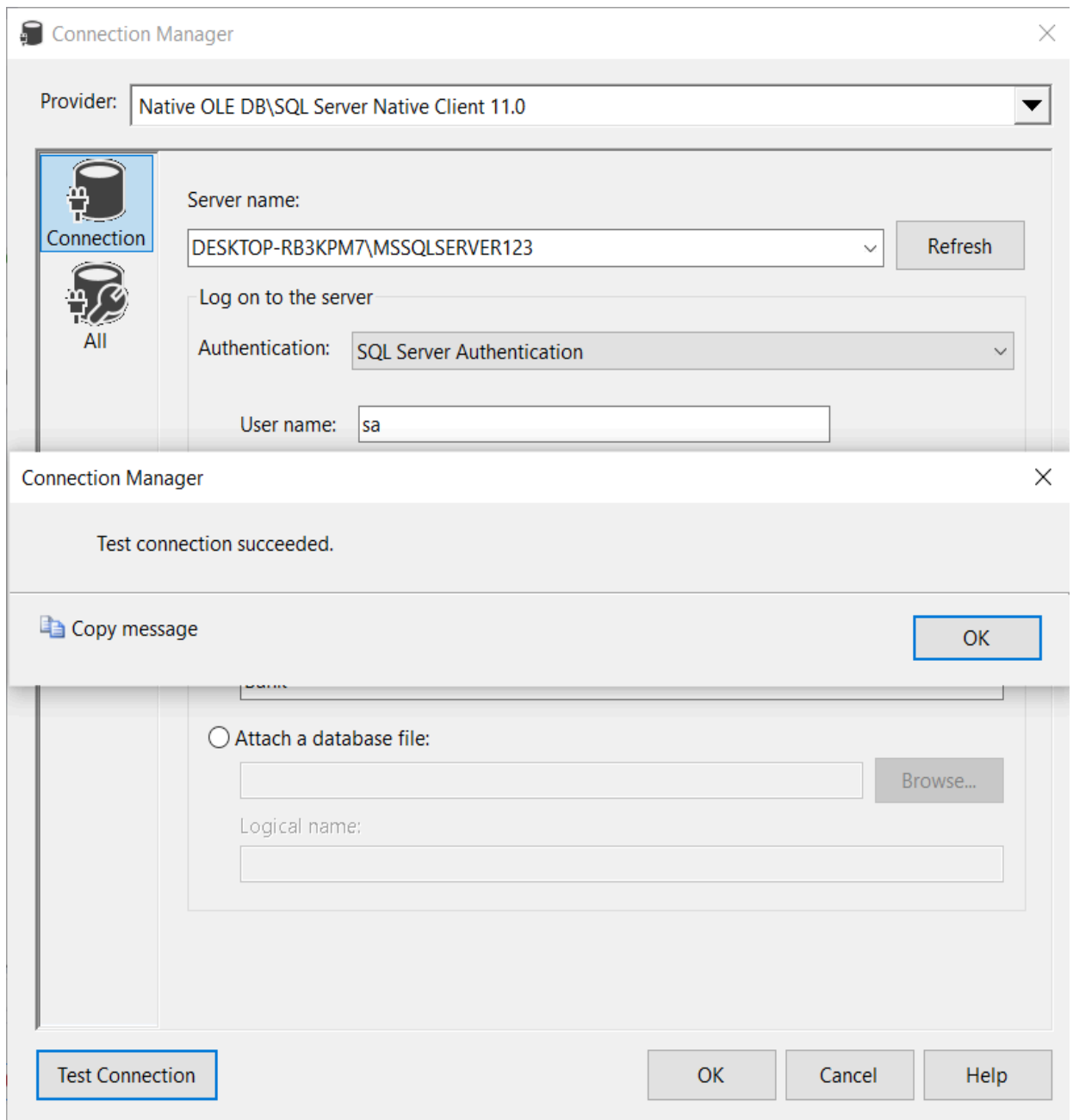
OLAP-куб дозволяє створювати багатовимірний аналіз даних. Процес його створення включає:

1. Створення проекту у Visual Studio:

- Відкрив Visual Studio і створив проект типу **Analysis Services Multidimensional and Data Mining Project**.
- Назвав проект

2. Налаштування джерела даних:

- Підключив базу даних **Bank** до проекту через Data Source.
- Використав облікові дані для підключення до SQL Server.

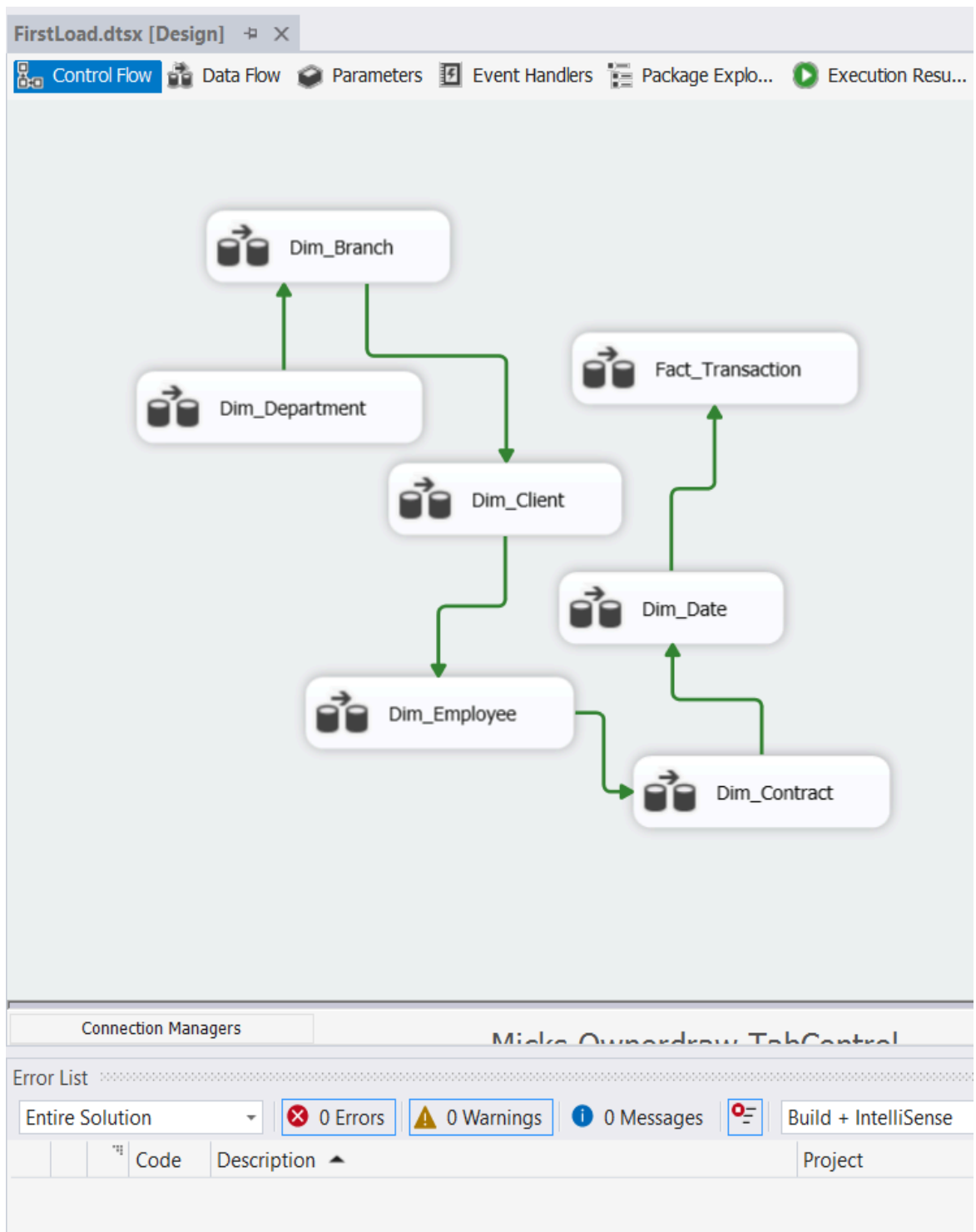


3. Створення Data Source View (DSV):

- Вибрав таблиці Clients і Transactions для включення до DSV.
- Підключив зв'язки вручну.

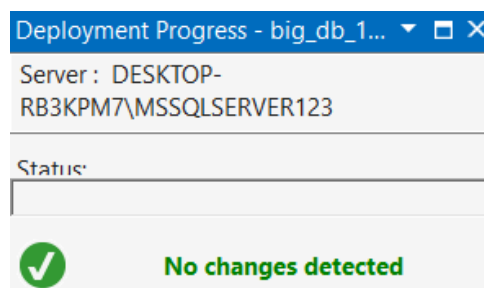
4. Налаштування мір:

- Додав міру **Сума транзакцій (Amount)**, яка обчислює загальну суму операцій.
- Додав міру **Середній кредитний рейтинг (CreditRating)** для аналізу рейтингу клієнтів.



5. Розгортання куба:

Розгорнув проект на сервері SSAS і перевірів його роботу через тестові запити MDX:



3.3. Генерація звітів

1. У Visual Studio:

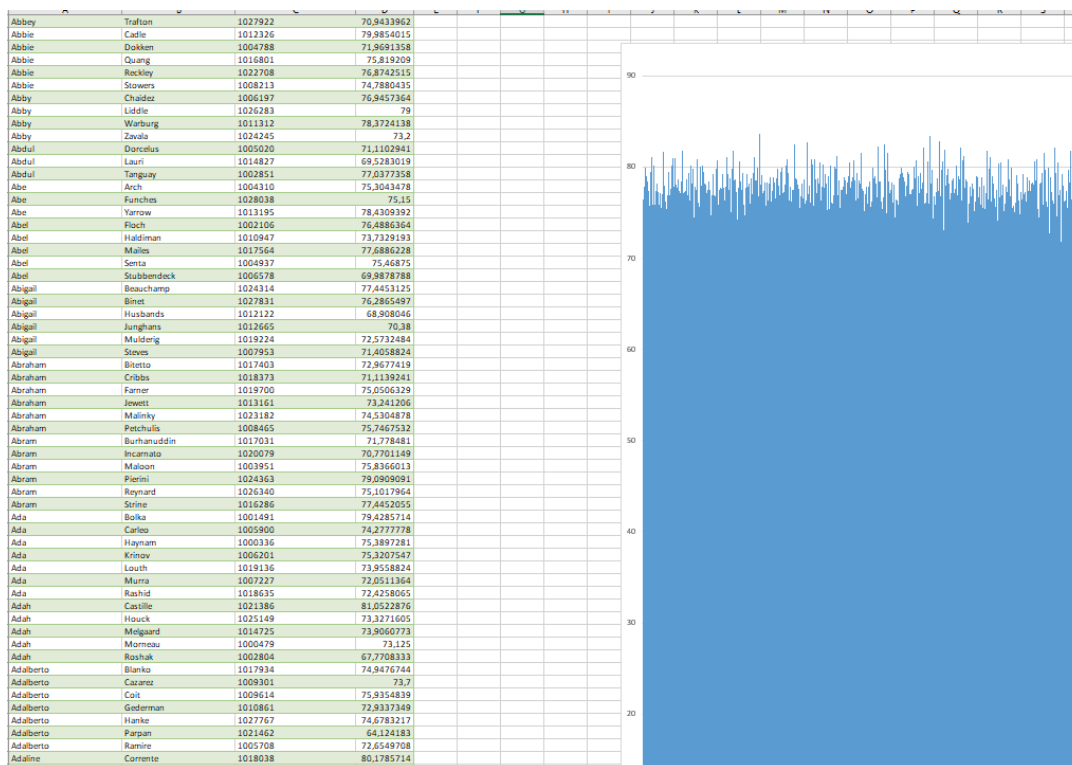
- Створив звіт, що відображає кредитний рейтинг клієнтів.
- Налаштував фільтри для сортування за часом або окремими клієнтами.

Report2.rdl [Design] ✕				
Design		Preview		
1	of 6 ?	100%	Find	
1000003	Tom	Schurz	6207	
1000004	Kasha	Villandry	70,2312925170068	
1000005	Jackqueline	Femia	69,4879518072289	
1000006	Brittani	Mannie	73,4666666666667	
1000007	Erica	Schurz	75,9537037037037	
1000008	Josefine	Mcmindes	75,3071895424837	
1000009	Yang	Guaman	75,7677419354839	
			72,35625	
1000011	Kraig	Franceschi	80,132183908046	
1000012	Illuminada	Blansett	76,0716417910448	
1000013	Kaleigh	Zukor	73,1032258064516	
1000014	Nathanial	Hornbarger	76,7456647398844	
1000015	Veta	Mcinture	73,5595238095238	
1000016	Rosy	Dingle	71,3575757575758	
1000017	Carmon	Roam	74,4810126582279	
1000018	Barney	Breedan	81,7658227848101	
1000019	Jorge	Staie	70,5816993464052	
1000020	Liberty	Polcyn	76,7361963190184	
1000021	Tony	Corav	73.514184397	

В Excel:

- Підключив OLAP-куб до Excel.

- Додав гістограму для візуалізації змін кредитного рейтингу клієнтів.



4. Висновки.

У ході виконання курсової роботи було досягнуто основної мети — розроблено аналітичне рішення для оцінки кредитного рейтингу клієнтів банку на основі їх транзакцій. Робота охопила всі необхідні етапи, починаючи від створення бази даних із нуля, проектування OLAP-куба, генерації звітів та закінчуючи аналізом отриманих даних.

Створення бази даних у SQL Server дозволило забезпечити надійну основу для подальшого аналізу. База даних включає дві основні таблиці — **Clients** та **Transactions**, які пов'язані між собою через унікальний ідентифікатор клієнта. Така структура дозволила забезпечити швидкий доступ до даних і зручність у використанні для побудови багатовимірної моделі.

Проектування OLAP-куба у Visual Studio дало змогу реалізувати багатовимірний аналіз даних. Використання вимірів, таких як **Час** та **Клієнти**, а також мір, зокрема **Сума транзакцій** і **Кредитний рейтинг**, дозволило отримати агреговані дані, що значно полегшило аналіз. Завдяки ієрархії виміру **Час** стало можливим вивчати динаміку змін кредитного рейтингу клієнтів у розрізі років, місяців та днів.

Генерація звітів у Visual Studio та Excel забезпечила зручну візуалізацію результатів. Звіти на основі OLAP-куба дозволили швидко визначити клієнтів із найвищим та найнижчим кредитним рейтингом, проаналізувати вплив транзакцій на їхній рейтинг, а також виділити тенденції та ризики у фінансовій діяльності клієнтів.

Аналіз даних виявив кілька важливих закономірностей. Наприклад, клієнти з активними депозитами мали вищий кредитний рейтинг, тоді як велика кількість зняття коштів негативно впливала на рейтинг. Такі висновки можуть бути використані для прийняття управлінських рішень щодо роботи з клієнтами, зокрема для надання кредитів чи визначення умов обслуговування.

Отже, результати роботи показали, що використання OLAP-кубів у поєднанні з сучасними інструментами, такими як SQL Server Management Studio та Visual Studio, є ефективним підходом до аналізу великих обсягів даних. Цей досвід є корисним для подальшого використання в реальних проектах і сприяє розвитку навичок роботи з аналітичними системами.

Джерела

1. Документація SQL Server: <https://docs.microsoft.com/sql/>
2. Документація Visual Studio: <https://learn.microsoft.com/visualstudio/>
3. Статті на тему OLAP: <https://www.sqlservercentral.com/articles/olap>
4. Документація MDX: <https://learn.microsoft.com/analysis-services/mdx>
5. Power BI Documentation: <https://learn.microsoft.com/power-bi/>
6. Книги з аналізу даних:
 - Kimball, R., Ross, M. "The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling".
 - Thomsen, E. "OLAP Solutions: Building Multidimensional Information Systems".
7. Практичні поради з роботи з Excel Pivot Tables: <https://support.microsoft.com/excel-pivot-tables>
8. Конспекти лекцій з курсу «Надвеликі бази даних»