

Lab 1 – Tasks 7 to 8 & Conclusion

Task 7: Deduplicate Data

Goal:

Remove duplicate loan purposes to clean the dataset.

Steps:

1. Identify duplicate values in the `purpose` column of the `loan` table.
2. Use the `DISTINCT` keyword to remove duplicates.

Key Questions & Answers:

- Which SQL keyword removes duplicates from query results?

 `DISTINCT`

- Challenge:

Create a table with distinct loan purposes:

SQL

```
CREATE TABLE fintech.loan_purposes AS
SELECT DISTINCT application.purpose AS purpose
FROM fintech.loan;
```

Afficher plus de lignes

Task 8: Answer Business Questions with a Report

Goal:

Generate a report showing the **total loan amount issued per year**.

Steps:

1. Identify relevant columns: `issue_year` and `loan_amount`
2. Use `GROUP BY` and `SUM()` to aggregate data:

SQL

```
SELECT issue_year, SUM(loan_amount) AS total_amount
FROM fintech.loan
GROUP BY issue_year;
```

Afficher plus de lignes

Key Questions & Answers:

- **Which columns store the data Trevor needs?**
☒ `issue_year` and `loan_amount`
- **Which SQL keyword creates groups in query results?**
☒ `GROUP BY`
- **What was the total amount of loans issued in 2019?**
☒ 852355425
- **Challenge:**
Create a table that counts loans per year:

SQL

```
CREATE TABLE fintech.loan_count_by_year AS  
SELECT issue_year, COUNT(loan_id) AS loan_count  
FROM fintech.loan  
GROUP BY issue_year;
```

Afficher plus de lignes

- **Which SQL keywords can you use to create a table?**
☒ `CREATE TABLE AS SELECT` (CTAS)

🏁 Conclusion

As a cloud data analyst at **TheLook Fintech**, you successfully:

- Explored and extracted key metrics from the loan dataset.
- Imported external data to enrich regional analysis.
- Joined tables to create meaningful reports.
- Deduplicated data to ensure clean results.
- Aggregated loan data by year to support cash flow analysis.

☒ You've demonstrated your ability to **collect**, **process**, and **store** data using **BigQuery**, and created **portfolio-ready outputs** that reflect real-world business needs.