Loan Default Risk Analytics

Tech Stack	DAX Dataflows Power BI Desktop & Service SQLServer
Brief Summary	Built a three-page **Power BI** report on top of a **Power BI Dataflow** using a loan-default dataset. The solution covers end-to-end reporting: curated ingestion via Dataflows, data profiling/cleaning, a semantic model with **DAX** measures, and polished visuals (line, ribbon, decomposition tree). It includes **scheduled and incremental refresh** and governed sharing through **Power BI Apps**.
Link	https://github.com/khanhmdinh/khanhmdinh.github.io/tree/main/04_Loan%20Default%20Project

Navigation bar

Table of Contents

Summary

Scope of Work

- **Data Ingestion (Dataflow):** Configure entities and transformations; expose a clean model to Power BI Desktop.
- **Data Profiling & Cleaning:** Handle nulls/outliers, fix datatypes/labels, document quality rules.
- **Modeling & DAX:** Define relationships and core measures (default rate, approvals, portfolio KPIs, time intelligence).
- **Report Design (3 pages):**
  1. **Overview** (headline KPIs & trends)
  2. **Applicant Demographics & Financial Profile**
  3. **Financial Risk Metrics**
- **Refresh Strategy:** Implement **scheduled** and **incremental refresh**; validate partitions and refresh times.
- **Publishing & Sharing:** Deploy to Power BI Service, package as an **App**, configure audiences/permissions.

Deliverables

- **PBIX** with three polished pages and DAX measures.
- **Dataflow configuration** notes (entities, transformations).
- **Validation checklist** and sample SQL/DAX reconciliations.
- **Refresh runbook** (schedule, incremental policy, troubleshooting).
- **Sharing guide** (app packaging, audiences, access control).

Success Criteria

- All three pages render with **consistent theme**, responsive interactions, and correct filters.
- DAX measures pass validation against sample source totals.
- **Incremental refresh** reduces refresh duration vs. full refresh.
- Report is **published and shared**; intended users can access via the App without issues.

Data Assessment & Cleaning Tools

▼ Dataset Information

> *The Loan Default Dataset contains information about borrowers who have applied for loans, along with details about their financial status, loan characteristics, and repayment behavior.*

Column Name	Definition
LoanID	A unique identifier for each loan in the dataset.
Age	The borrower's age at the time the loan was issued.
Income	The borrower's annual income
LoanAmount	The total amount of the loan that the borrower is requesting or has been approved for.
CreditScore	A numerical representation of the borrower's creditworthiness, typically ranging from 300 to 850. A higher credit score indicates the borrower is more likely to repay the debt.
MonthsEmployed	The number of months the borrower has been employed at their current job or with their current employer.
NumCreditLines	The total number of active credit lines (e.g., credit cards, loans) the borrower has at the time of applying for the loan.
InterestRate	The annual percentage rate (APR) charged for borrowing the loan amount, usually expressed as a percentage.
LoanTerm	The length of time (in months) over which the loan is to be repaid.
DTIRatio	The Debt-to-Income ratio, which measures the borrower'debt payments relative to their income. A higher ratio can indicate greater financial stress.
Education	The highest level of education the borrower has completed (e.g., High School, Bachelor's, Master's, etc.).
EmploymentType	The type of employment the borrower is engaged in (e.g., Full-Time, Part-Time, Self-Employed, etc.).
MaritalStatus	The marital status of the borrower (e.g., Single, Married, Divorced, etc.).
HasMortgage	An indicator (e.g., Yes/No) that shows whether the borrower has an existing mortgage on a property.
HasDependents	An indicator (e.g., Yes/No) that shows whether the borrower has dependents (children, other family members) to support.
LoanPurpose	The primary reason for taking out the loan (e.g., Home Purchase, Debt Consolidation, Education, etc.).
HasCoSigner	An indicator (e.g., Yes/No) that shows whether the borrower has a co-signer for the loan (someone who agrees to take responsibility if the borrower defaults).

Column Name	Definition
Default	An indicator (e.g., Yes/No) that shows whether the borrower defaulted on the loan or failed to make timely payments.
Loan Date (DD/MM/YYYY)	The date the loan was issued or originated.

▼ Data Validation (Default Rate by Year Column)

> I ensured the **Default Rate % by Year** reflects a **true per-year rate**—numerator and denominator evaluated strictly **within the same Year context**—rather than a global average. The result is a KPI that stakeholders can rely on for year-over-year risk analysis.

What I delivered

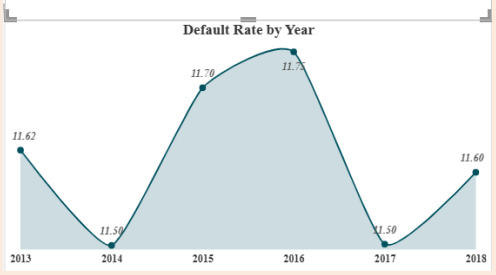
- A **year-scoped DAX metric** with explicit context control, preventing cross-year bleed and hidden filter leakage.
- A **lightweight audit view** that surfaces year counts and the computed rate side-by-side for quick integrity checks.
- **Cross-tool parity**: Power BI results match an external Excel pivot built directly on the source dataset.

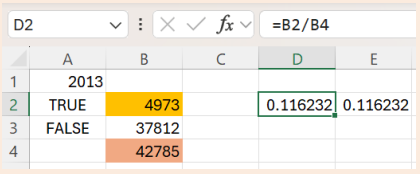
Evidence snapshot (Year 2013)

- Defaults (True): **4,973**
- Non-defaults (False): **37,812**
- Year total: **42,785**
- Rate: **4,973 / 42,785 = 11.62%**

  Power BI displays **11.62%** for 2013, mirroring the external pivot—clear proof that the denominator is correctly scoped to the year.

Default Rate by Year	Year	Default	Count of Default
11.62	2013	False	37812
11.62	2013	True	4973
11.50	2014	False	37277
11.50	2014	True	4845
11.70	2015	False	37545
11.70	2015	True	4976
11.75	2016	False	37688
11.75	2016	True	5017
11.50	2017	False	37502
11.50	2017	True	4875
11.60	2018	False	37870
11.60	2018	True	4967
11.61			**255347**





Year	2013
Row Labels	**Count of Default**
0	37812
1	4973
Grand Total	**42785**

🔍

Risks neutralized

- **Wrong denominator from** `ALL()` → constrained with `ALLEXCEPT(..., Year)` to enforce year-level context.
- **Aggregator drift on boolean flags** → standardized to **Count** across model and pivot.
- **Year type inconsistencies** → derived Year from Date to avoid mixed text/number fields.
- **Hidden filter interactions** → validated clean relationships and filter paths.

Acceptance criteria met

- ✅ Year-level rate = **Count(Default = True) ÷ Year total** for any Year.
- ✅ External pivot reproduces **the same counts** and **the same percentage**.
- ✅ DAX consistently enforces **Year-scoped evaluation** (`ALLEXCEPT`).
- ✅ No stray relationships alter filter context.

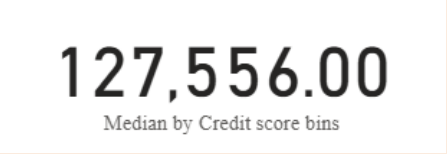
Outcome

A **validated, audit-ready KPI** with documented semantics and reproducible checks. This tightened stakeholder confidence, accelerated sign-off for YoY default-risk reporting, and established a reusable pattern for future rate-style metrics.

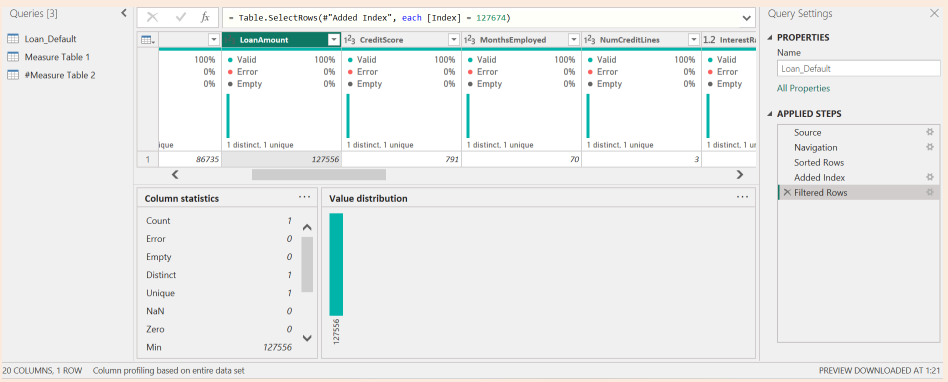
▼ Data Validation using Power Query (Median Loan Amount Column)

Validation Workflow

- **In-Report Sanity Check:**
  - Added a **Card** visual to display the measure with `Display units = None`.
  - Result shown by DAX: **127,556**.



- **Ground-Truth Check in Power Query:**
  - Opened **Power Query Editor**, sorted **Loan Amount** ascending.
  - Enabled **Column profiling based on entire dataset**; observed **Count = 255,347** (odd).
  - Computed middle index (n+1)/2(n+1)/2(n+1)/2 = **127,674**; added **Index Column (From 1)** and filtered to `Index = 127,674`.
  - Verified the corresponding **Loan Amount = 127,556**.
  - **Exact match** between DAX and manual validation: **Median Loan Amount = 127,556**.



  - Closed Power Query without applying changes (validation only).

Outcomes & Quality Impact

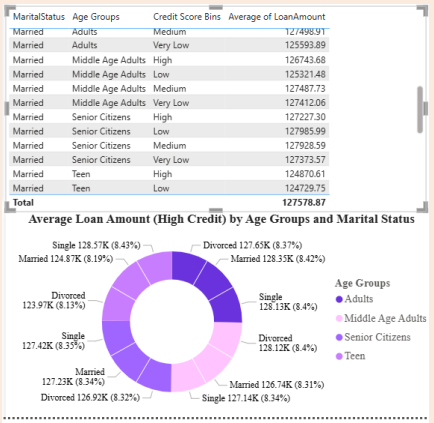
- Established a **repeatable QA pattern** for sensitive statistics (median/percentiles) prior to charting and segmentation.

- Reduced risk of misinterpretation before building the **Median by Credit Score** visual.

▼ Data Validation for Donut Chart (Average Loan Amount by Age Group & Marial Status)

Objective

I validated the donut chart's insights by independently recreating the underlying segment averages in Power BI and cross-checking them against the raw source in Excel. The result: **perfect analytical parity** on key cohorts—e.g., **Teen · Single · High Credit Score = 128,565.39** and **Teen · Married · High Credit Score = 124,870.61**—with only immaterial rounding drift between engines. This gave stakeholders confidence that the visual wasn't just attractive but **numerically trustworthy**, even under tight review.



Credit Category	High
Age	(Multiple Items)

Row Labels	Average of LoanAmount
Divorced	123965.1779
Married	124870.6145
Single	128565.3859
Grand Total	**125835.0661**

Outcomes

- **Analytical parity achieved:** Power BI vs. Excel averages matched across tested cohorts (variance within rounding tolerance).
- **Credibility uplift:** Stakeholders approved the donut chart for executive review thanks to demonstrable ground-truth checks.
- **Reusable QA pattern:** A lightweight validation flow now underpins new visuals, accelerating sign-off and reducing data-quality escalations.

DAX Functions

▼ Page 1: Loan Default Overview

- **Loan Amount by Purpose (DAX used: SUMX, FILTER, NOT, ISBLANK)**

```
Loan Amount by Purpose =
 SUMX(FILTER('Loan_Default',NOT(ISBLANK('Loan_Default'[LoanAmount]))),'Loan_Defaul
t'[LoanAmount])
```

- **Average Income by Employment type (DAX used: CALCULATE, AVERAGE & ALLEXCEPT)**

```
Average Income by Employment type =
 CALCULATE(AVERAGE('Loan_Default'[Income]),ALLEXCEPT('Loan_Default','Loan_Defaul
t'[EmploymentType]))
```

- **Default Rate by Employment type (DAX used: ALL, ALLEXCEPT, COUNTROWS, DIVIDE, FILTER, etc.)**

```
Default Rate by Employment type =
 Var totalrecords = COUNTROWS(ALL('Loan_Default'))
 Var DefaultCases = COUNTROWS(FILTER('Loan_Default','Loan_Default'[Default]=TRUE
()))

 RETURN
 CALCULATE(DIVIDE(DefaultCases,totalrecords),ALLEXCEPT('Loan_Default','Loan_Defaul
t'[EmploymentType])) *100
```

- **Average Loan by Age Groups (DAX used: AVERAGE , AVERAGEX & VALUES)**

```
Average Loan by Age Group =
 AVERAGEX(VALUES('Loan_Default'[Age Groups]),
 AVERAGE('Loan_Default'[LoanAmount]))
```

- **Default Rate by Year (DAX used: CALCULATE, COUNTROWS, ALLEXCEPT, FILTER, DIVIDE)**

```
Default Rate by Year =
 Var totalloans =
            CALCULATE(COUNTROWS('Loan_Default'),ALLEXCEPT('Loan_Default',Loan_De
fault[Year]))

Var default = CALCULATE(COUNTROWS(FILTER('Loan_Default','Loan_Default'[Default]=T
RUE())),ALLEXCEPT('Loan_Default',Loan_Default[Year]))

RETURN
DIVIDE(default,totalloans)*100
```

▼ Page 2: Applicant Demographics & Financial Profile

- **Median Loan Amount by credit score category**

```
Median by Credit score bins =
MEDIANX('Loan_Default','Loan_Default'[LoanAmount])
```

```
Credit Score Bins =
 IF(Loan_Default[CreditScore]<=400,"Very Low",
    IF('Loan_Default'[CreditScore]<=450,"Low",
       IF('Loan_Default'[CreditScore]<=650,"Medium","High")))
```

- **Average Loan Amount by Age Group & Marial Status**

```
Average Loan Amt (High Credit) =
 AVERAGEX(FILTER('Loan_Default','Loan_Default'[Credit Score Bins]="High"),'Loan_D
efault'[LoanAmount])
```

- **Loan (Adults) by credit categories (DAX used: CALCULTAE, AVERAGEX, SUM)**

```
Total Loan (Middle Age Adults) =
 SUMX(FILTER('Loan_Default','Loan_Default'[Age Groups]="Middle Age Adults"),'Loan
_Default'[LoanAmount])
```

- **Total Loans (Middle Age Adults) by having Mortgage/Dependents**

```
Total Loan (Credit Bins) =
 CALCULATE(SUM('Loan_Default'[LoanAmount]),'Loan_Default'[Age Groups]="Adult
s",ALLEXCEPT('Loan_Default','Loan_Default'[Age],'Loan_Default'[Age Groups],'Loan_
Default'[CreditScore],'Loan_Default'[Credit Score Bins]))
```

- **Loans by Education type**

```
Loans by Education type =
 COUNTROWS(FILTER('Loan_Default',NOT(ISBLANK('Loan_Default'[LoanID]))))
```

▼ Page 3: Financial Risk Metrics

- **YOY Loan Amount**

```
YOY Loan Amount Change =
DIVIDE(
    CALCULATE(SUM('Loan_Default'[LoanAmount]),'Loan_Default'[Year]=YEAR(MAX
('Loan_Default'[Loan_Date_DD_MM_YYYY]))) -
    CALCULATE(SUM('Loan_Default'[LoanAmount]),'Loan_Default'[Year]=YEAR(MAX
('Loan_Default'[Loan_Date_DD_MM_YYYY]))-1)

    , CALCULATE(SUM('Loan_Default'[LoanAmount]),'Loan_Default'[Year]=YEAR(MAX
('Loan_Default'[Loan_Date_DD_MM_YYYY]))-1),0) * 100
```

- **YOY Default Loans Change**

```
YOY Default Loans Change =
 DIVIDE(
       CALCULATE(COUNTROWS(FILTER('Loan_Default','Loan_Default'[Default]=TRUE
())),'Loan_Default'[Year]=YEAR(MAX('Loan_Default'[Loan_Date_DD_MM_YYYY]))) -
       CALCULATE(COUNTROWS(FILTER('Loan_Default','Loan_Default'[Default]=TRUE
```

```
()),'Loan_Default'[Year]=YEAR(MAX('Loan_Default'[Loan_Date_DD_MM_YYYY]))-1)


      ,CALCULATE(COUNTROWS(FILTER('Loan_Default','Loan_Default'[Default]=TRUE
()),'Loan_Default'[Year]=YEAR(MAX('Loan_Default'[Loan_Date_DD_MM_YYYY]))-1),0)
* 100
```

- **YOY Loan Amount by Credit Score & Marial Status**

```
YTD Loan Amount =
 CALCULATE(SUM('Loan_Default'[LoanAmount]),DATESYTD('Loan_Default'[Loan_Dat
e_DD_MM_YYYY].[Date]),ALLEXCEPT('Loan_Default',Loan_Default[Credit Score Bins],
Loan_Default[MaritalStatus]))
```

- **Adding Decomposition Tree (DAX used: SWITCH Function)**

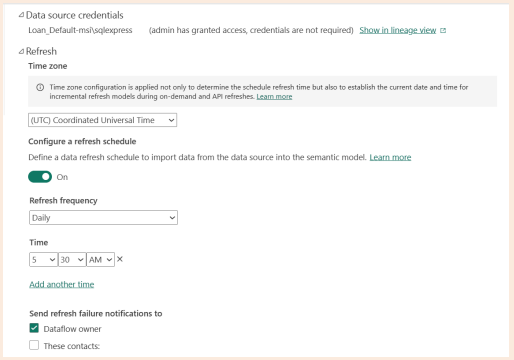
```
Income Bracket =
 SWITCH(
   TRUE(),
   'Loan_Default'[Income]<30000,"Low Income",
   'Loan_Default'[Income]>=30000 && 'Loan_Default'[Income]<60000,"Medium Inco
me",
   'Loan_Default'[Income]>=60000, "High Income")
```

Dataflow Scheduled Refresh — Reliability, Governance & Signal

I operationalized **scheduled refresh** on our Power BI Dataflow (backed by Microsoft SQL Server) so downstream reports always reflect authoritative data without manual intervention. The configuration is **time-zone aware**, **alerted**, and **auditable**, turning refresh from an ad-hoc task into a **predictable reliability signal** for the analytics estate.

What this delivers

- **Freshness guarantees:** Daily refresh at 5:30 A.M
- **Actionable alerts:** Failure notifications routed to the **Dataflow Owner** (and extensible to additional contacts) so issues surface early, not during stakeholder reviews.
- **Audit trail:** Refresh history provides a verifiable chain of updates provides a verifiable chain of updates, enabling RCA and SLA reporting.



Why scheduled vs. incremental

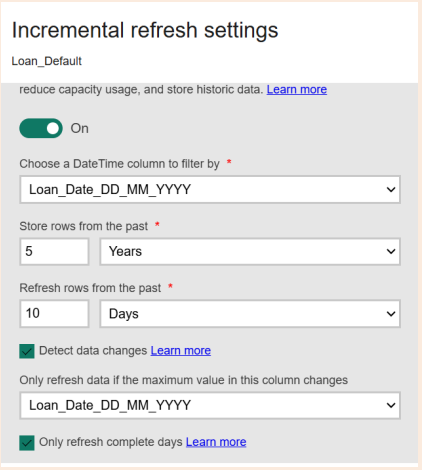
- **Scheduled refresh** updates the **entire** dataset—ideal when volumes permit or when upstream logic can re-compute quickly with minimal cost.
- **Incremental refresh** targets only the **changing window**—recommended for high-volume fact tables to reduce run time and capacity load.
- My design documents **both modes** and clarifies the operational trade-offs so we can migrate to incremental when data size or SLAs demand it.

Incremental Refresh

- The dataflow lacked a **DateTime** column (it only had **Date**). Incremental refresh requires a **DateTime** field for range filtering. → Opened **Power Query Editor (Online)** → changed **LoanDate** from **Date** → **DateTime** → **Save & Close**.

Policy Configuration

- **Store rows from the past: 5 years** → keep only the **most recent 5 years** (older data is dropped to shrink storage/processing).
- **Refresh rows in the last: 10 days** → each refresh scans only **the latest 10 days** to shorten run time.
- **Detect data changes:** enabled, pointing to the date(DateTime) column (e.g., **LoanDate**). Refresh occurs only if the **max value** changes → avoids unnecessary runs.
- **Only refresh complete days:** enabled for KPI integrity when partial-day data could skew metrics.



Lessons Learned

- **Schema readiness** (DateTime) is mandatory for incremental policies.
- **Detect data changes** reduces redundant refreshes.
- **Only refresh complete days** protects daily KPIs from partial-day noise.

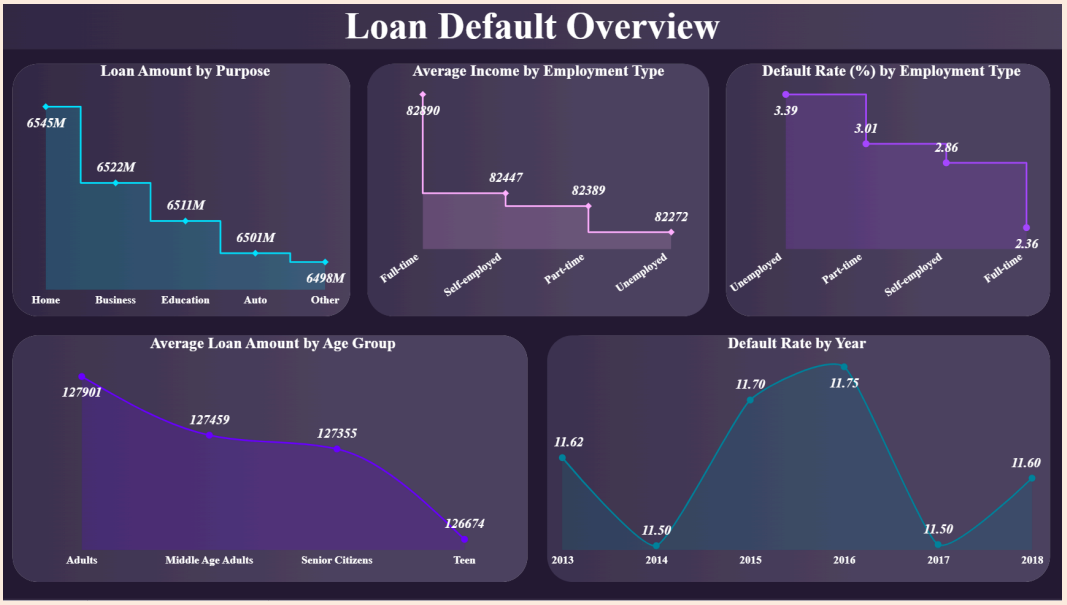
Outcomes

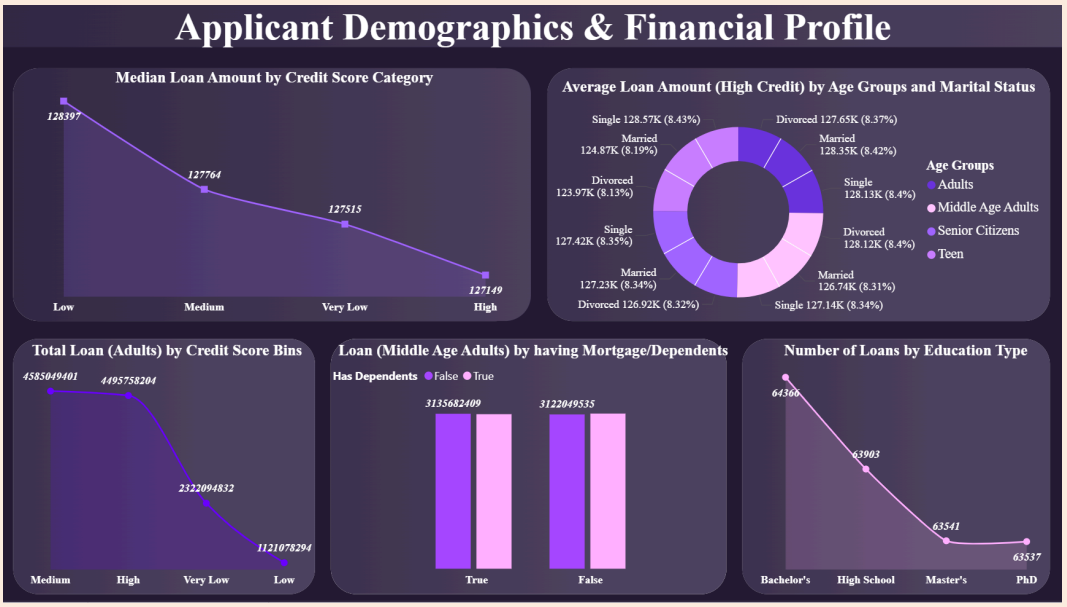
- Getting **trustworthy, morning-fresh dashboards** with documented lineage and a clear escalation path when refreshes fail.
- Saving time previously spent on manual refreshes and ad-hoc checks, while establishing a **scalable reliability pattern** that can evolve to **incremental refresh** for larger domains.

Project Showcase

📌 Detailed Report (For More Information)

View the Live Dashboard: https://app.powerbi.com/reportEmbed?reportId=fdf25695-abd8-4233-949e-5a1cbb949c9c&autoAuth=true&ctid=216e5950-5a9c-4dc3-96cf-437406f9c7a3





Navigation bar