Loan defaulters

A summary for a bank dataset

What is meant by loan defaulter?

Default is the failure to repay a loan. A default can occur when a borrower is unable to make timely payments, misses payments, or avoids or stops making payments.

When an applicant requests a loan from a certain bank, the bank is subjected to two types of risks:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the bank

2. If the applicant is not likely to repay the loan, then approving the loan may lead to a financial loss for the bank.

So our aim is to provide insightful dashboards for the bank about what characteristics are often associated with loan defaulters.

This will ensure that future loan decisions of the bank are made more logically and reduce possible defaults!

The Project was done using SQL Server, SSDT(SSIS, SSAS), and Power BI.

Stakeholders:

- Lender: Bank
- borrower: Clients
- Loan department

Table of Contents

01

Phase one

Business overview, Extracting data, Data cleaning, and transformations:

- Extracting data in the form of CSV files.
- Exploring and understanding the dataset.
- Removed columns with more than 40% of missing values.
- Transforming negative values into positive values.
- Transforming columns that has duration as days to duration in years.

Business overview, Extracting data, Data cleaning, and transformations:

- Added new columns that has data categories.
- Renaming columns to meaningful headers.
- Added “date” dimension.
- Applied aggregate functions on some columns to deliver more meaningful data form.
- Change DataType.

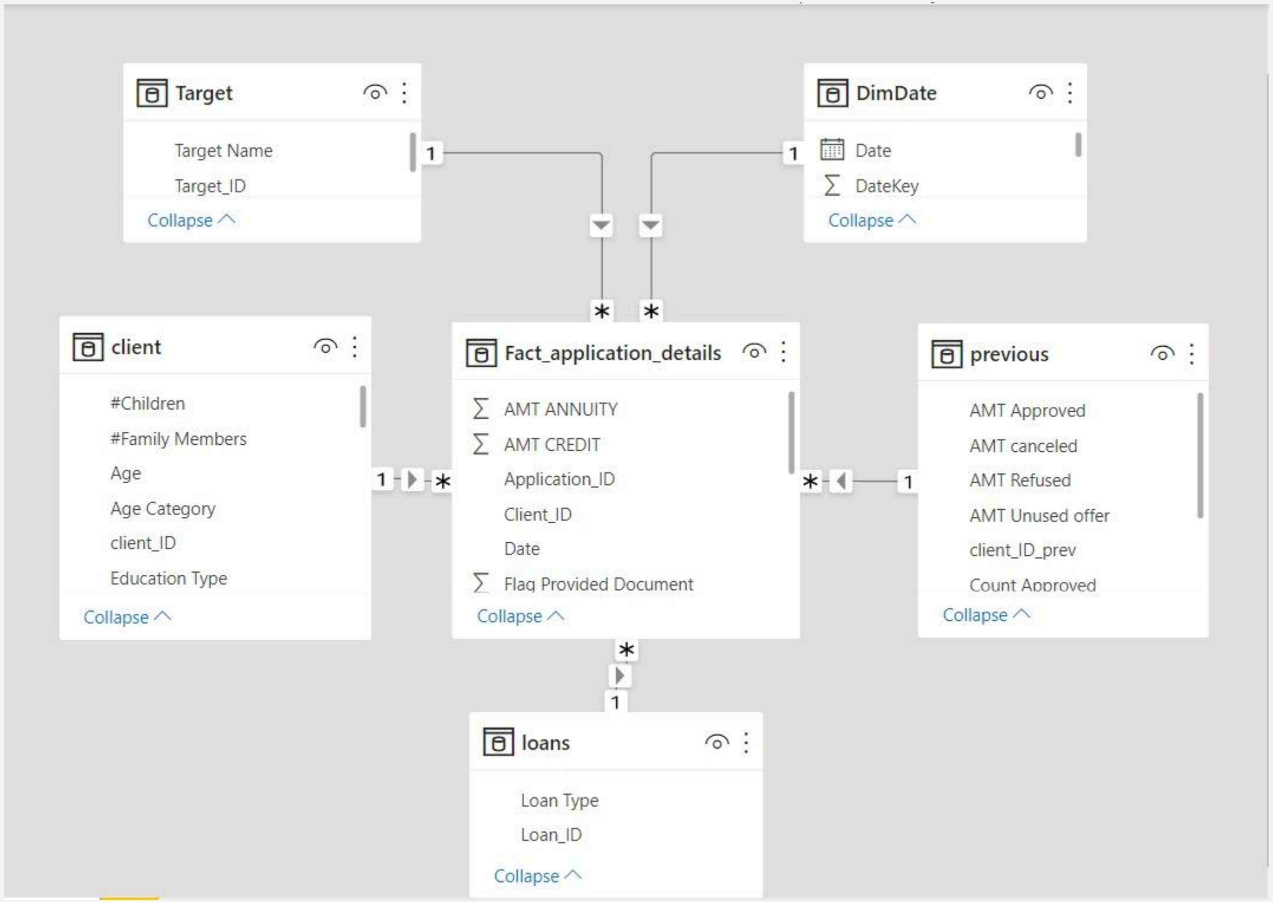
02

Phase two

Load data into a Data warehouse:

Using SSIS to transform data from CSV to a database form and create the fact table of the star schema.

Star Schema



Target
Target Name
Target_ID
Collapse
DimDate
Date
DateKey
Collapse
client
#Children
#Family Members
Age
Age Category
client_ID
Education Type
Collapse
Fact_application_details
AMT ANNUITY
AMT CREDIT
Application_ID
Client_ID
Date
Flag Provided Document
Collapse
previous
AMT Approved
AMT canceled
AMT Refused
AMT Unused offer
client_ID_prev
Count Approved
Collapse
loans
Loan Type
Loan_ID
Collapse
1
1
1
1
1
*
*
*
*
*

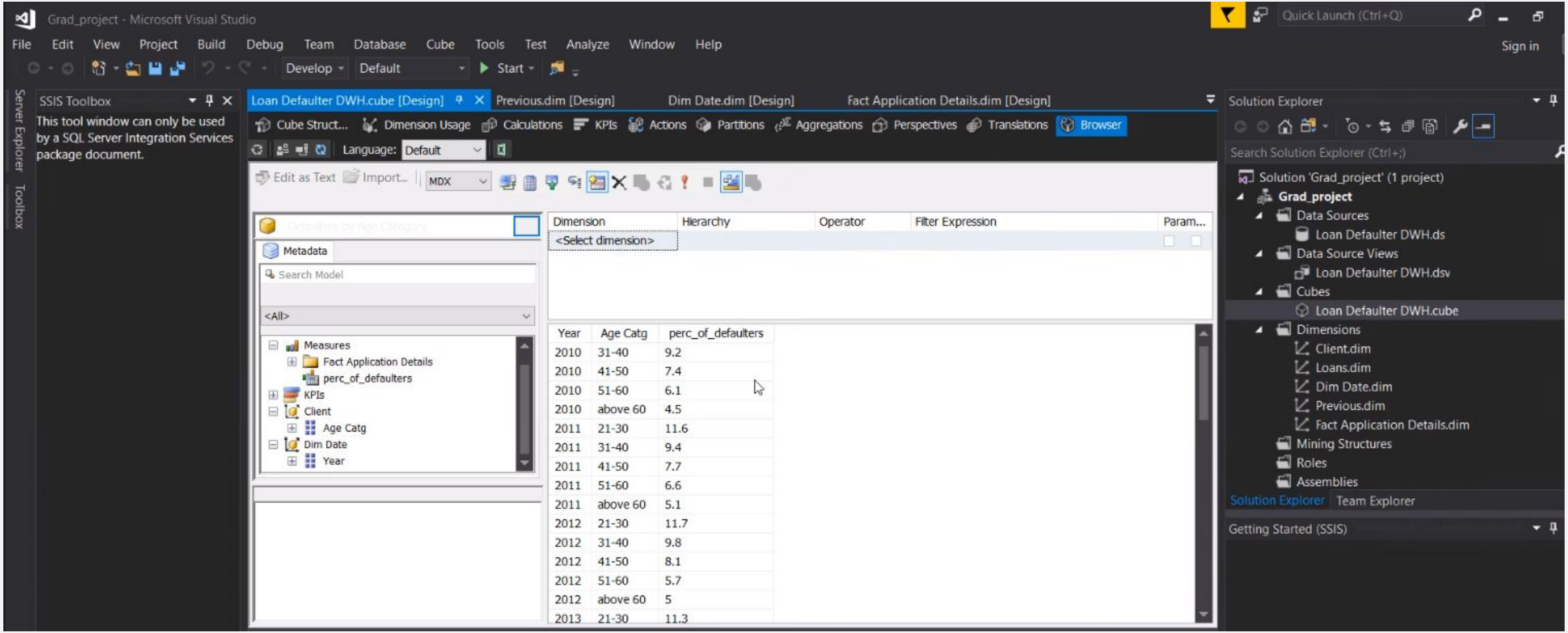
03 Phase three

Creating cubes, Perspectives, and KPI by SSAS:

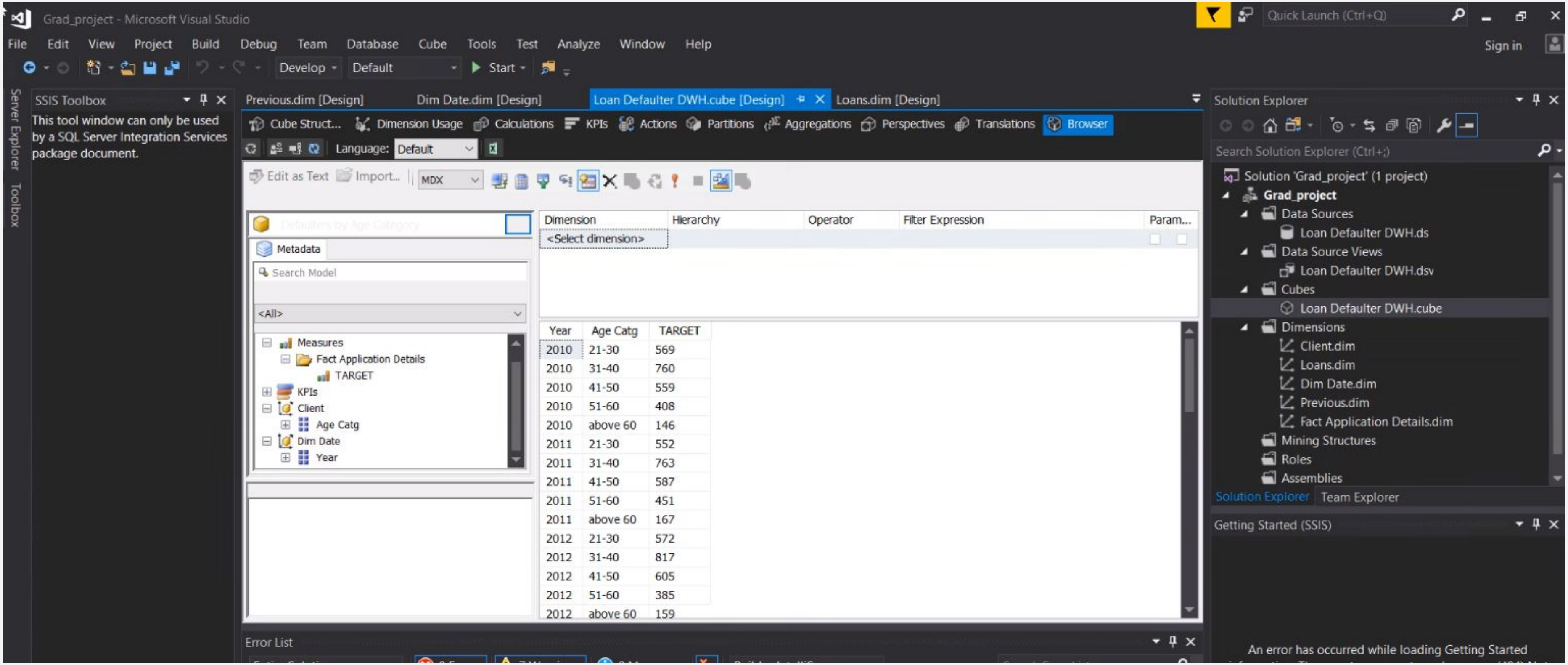
Created a cube that has 3 perspectives.

- **Percentage of defaulters perspective**
- **#Defaulters by age category perspective**
- **Avg credit by income type perspective**
- **Risk KPI**

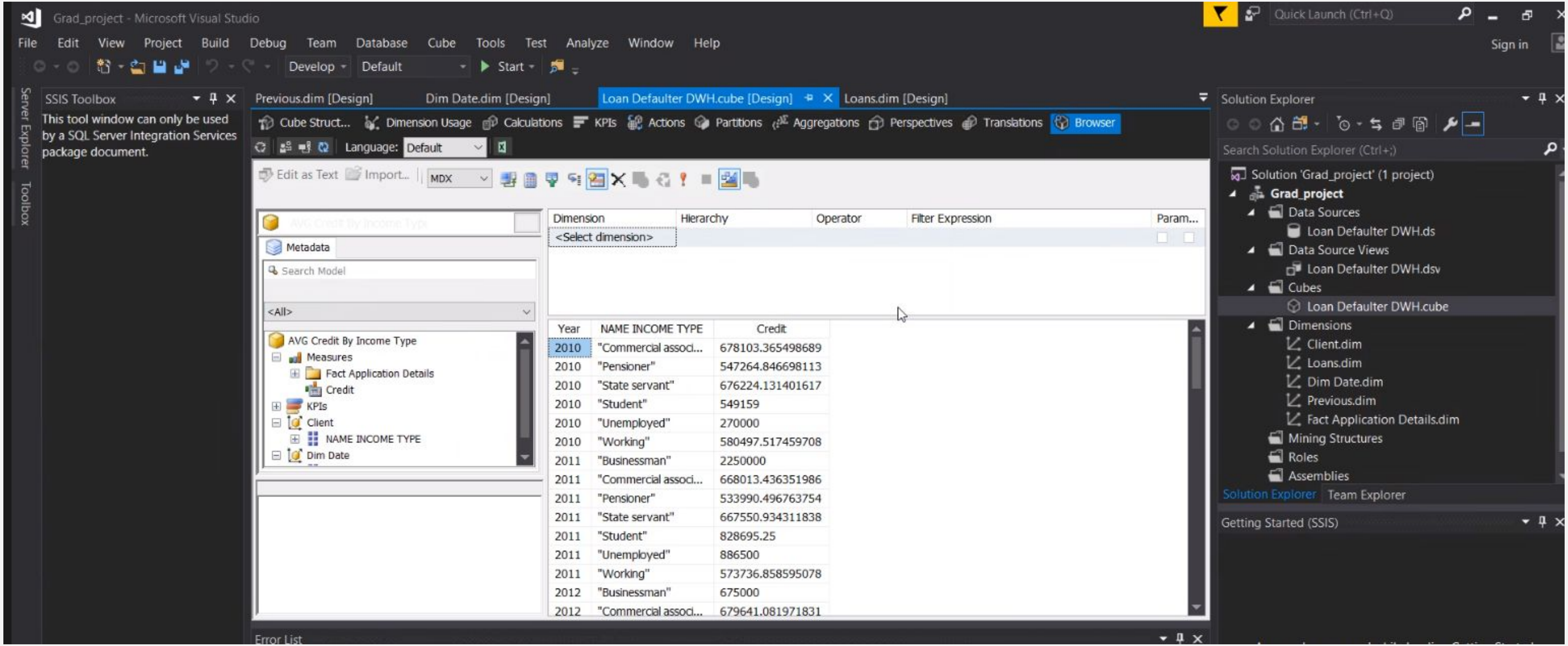
Percentage of defaulters perspective



#Defaulters by age category perspective



Avg credit by income type perspective



Risk KPI

Row Labels	TARGET	percentage	Risk Status
2010			
21-30	569	11.7	
31-40	760	9.2	
41-50	559	7.4	
51-60	408	6.1	
above 60	146	4.5	
2011			
21-30	552	11.6	
31-40	763	9.4	
41-50	587	7.7	
51-60	451	6.6	
above 60	167	5.1	
2012			
21-30	572	11.7	
31-40	817	9.8	
41-50	605	8.1	
51-60	385	5.7	
above 60	159	5	
2013			
21-30	544	11.3	
31-40	785	9.5	
41-50	554	7.4	
51-60	396	5.9	
above 60	157	4.9	
2014			
21-30	559	11.5	
31-40	805	9.6	
41-50	581	7.6	
51-60	406	6	
above 60	178	5.5	

04

Phase four

Data visualization with interactive dashboards using Power BI:

Loan information insights

- The percentage of defaulters in Cash loans is higher than that of Revolving loans.
- The number of defaulters in loan applications formed 25K defaulters of 308K total applications.
- Total Applications is 308k.
- Number of non-defaulters formed 283K of 308k total applications.
- Only 8.07% were defaulters.
- Females submit more loan applications than males.
- 2016 witnessed highest number of loan applications.
- 2012 witnessed highest number of loan defaulters.

Defaulters personal information

- Number of female defaulters is more than the number of male defaulters.
  yet, the females are more likely to repay the loan and not default.
- The age category 31-40 is more likely to default.
- Married applicants are more likely to default.
- The higher the number of family members

Professional information

- Applicants who are workers are more likely to default.
- The Employment duration is inversely proportional to default rates.
- Applicants with business entity type 3 (Transport) are more likely to default.
- laborers showed the highest rate of defaulting.

Financial information

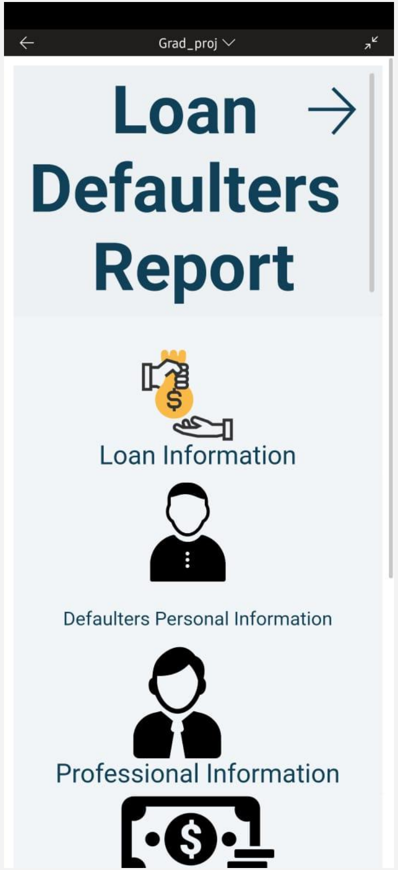
- Applicants with higher incomes are less likely to default.
- Most of the loan applications submitted are by people whose housing type is House/apartment.
- Applicants with housing types (rented apartments and with parents) are more likely to default.
- Applicants who don’t own cars submits more loan applications.
- Applicants who own cars are less likely to default.

Previous loans for current clients and risk information

- Applicants with more than one previous loan are less likely to default except for clients with more than 7 previously approved loans.
- Applicants who submitted Document_3 are less likely to default.
- Applicants with region rating 3 are more likely to default.
- Applicants whose living city is not the same as their working city are more likely to default.
- The percentage of defaulters in applicants whose ID is recently published is high.

05 Phase five

Creating a mobile layout for the project:




Phase six
06

Publishing the project on the Power BI service:

Thank you