

A blue-tinted photograph of the New York Stock Exchange building facade, featuring large classical columns and the words 'NEW YORK STOCK EXCHANGE' carved into the stone above them.

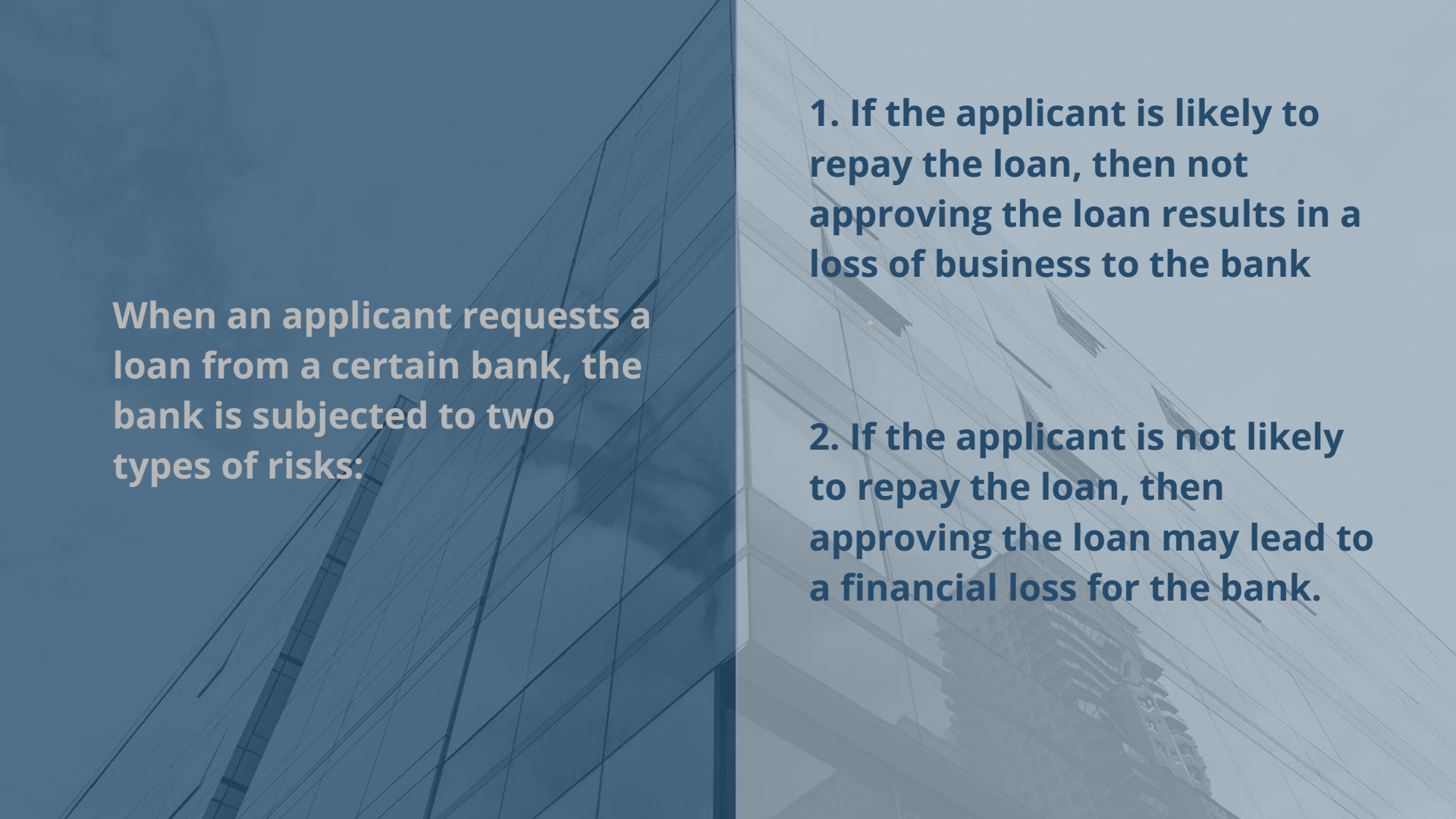
NEW YORK STOCK EXCHANGE

# 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 transformation.

02

## Phase two

Load data into a Data warehouse.

03

## Phase three

Creating cubes, Perspective, and KPI by SSAS.

04

## Phase four

Data visualization with interactive dashboards using Power BI.

05

## Phase five

Creating a mobile layout for the project.

06

## Phase six

Publishing the project on the Power BI service.

A low-angle, upward-looking photograph of several modern skyscrapers with glass facades. The buildings are set against a clear blue sky with a few wispy clouds. A white rectangular frame is superimposed over the center of the image, enclosing the text.

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.**



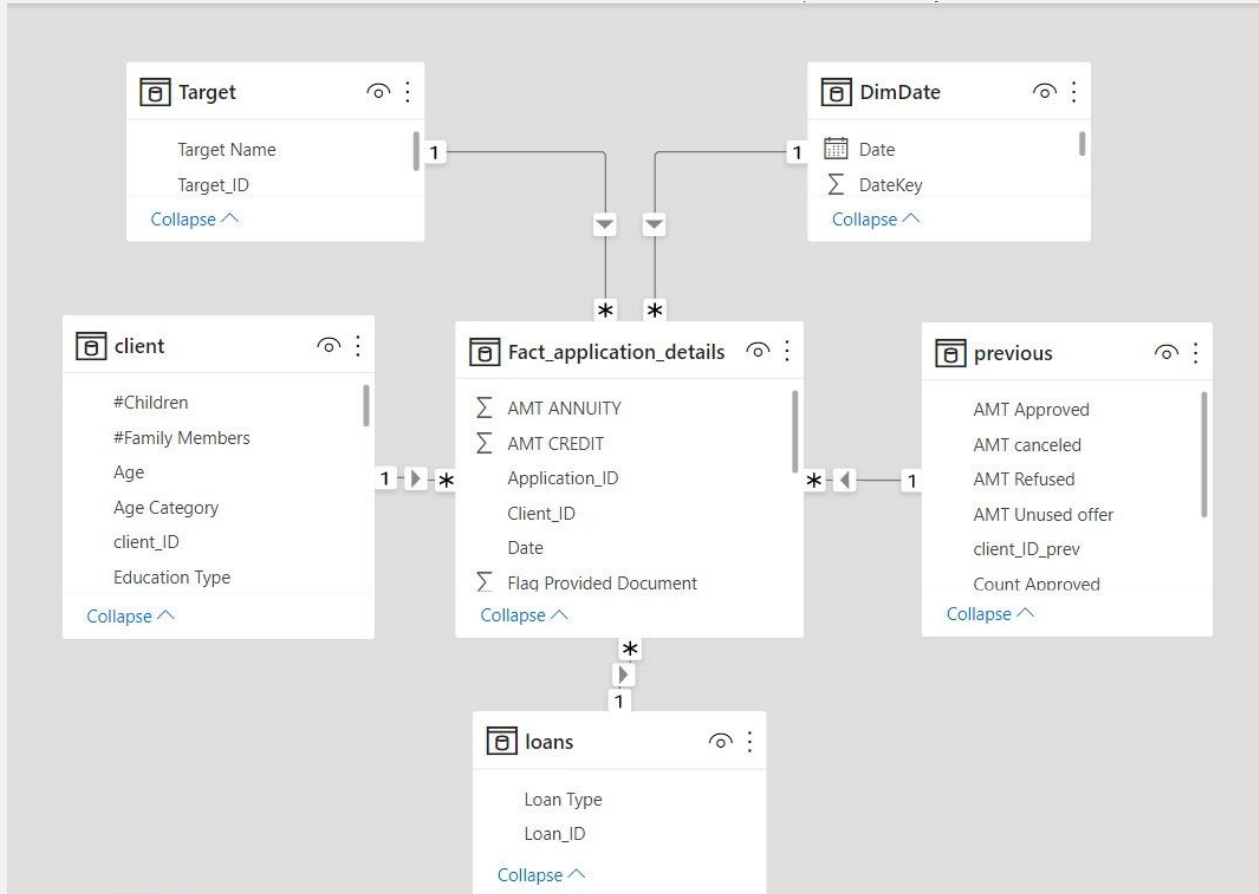
Phase two

02

**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



A low-angle, upward-looking photograph of several modern skyscrapers with glass facades, set against a blue sky with scattered white clouds. The perspective creates a sense of height and scale. A white rectangular frame is superimposed over the center of the image, containing the text.

03

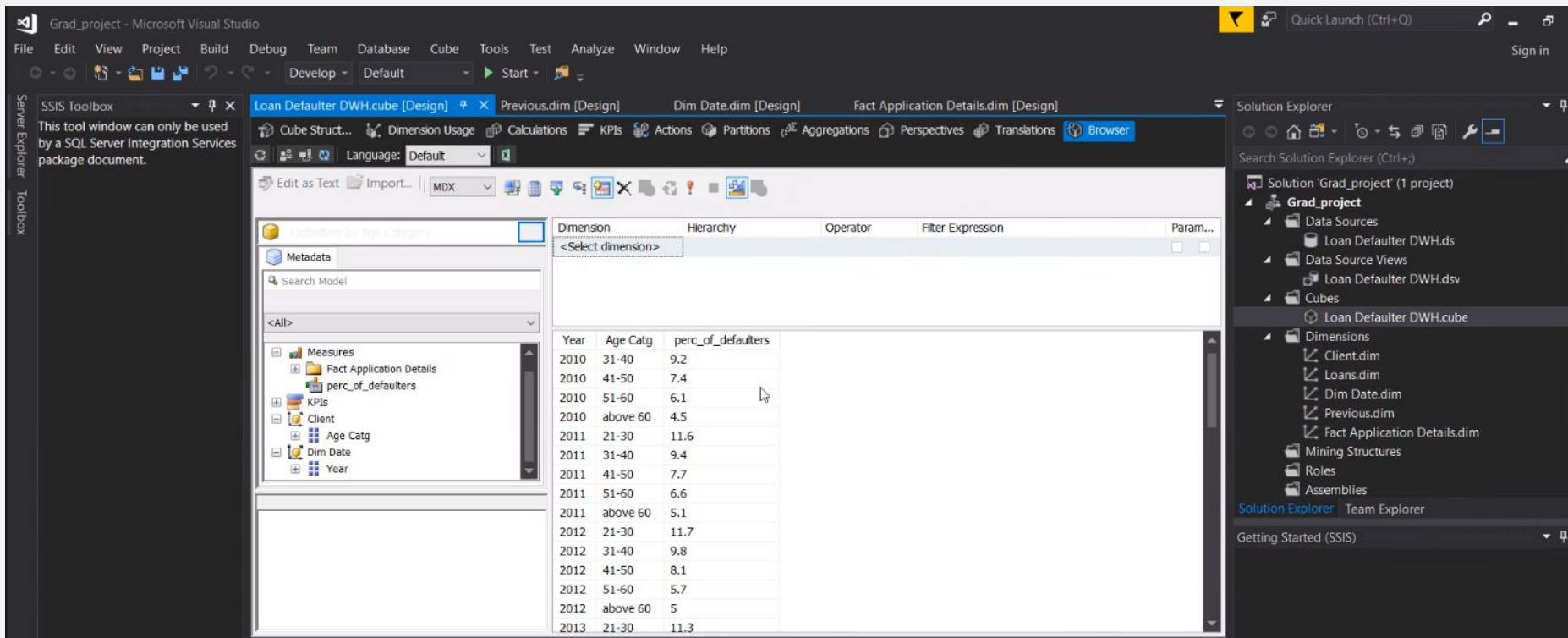
Phase three

## 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



Grad\_project - Microsoft Visual Studio

File Edit View Project Build Debug Team Database Cube Tools Test Analyze Window Help

Develop Default Start

SSIS Toolbox

This tool window can only be used by a SQL Server Integration Services package document.

Loan Defaulter DWH.cube [Design] Previous.dim [Design] Dim Date.dim [Design] Fact Application Details.dim [Design]

Cube Struct... Dimension Usage Calculations KPIs Actions Partitions Aggregations Perspectives Translations Browser

Language: Default

Edit as Text Import... MDX

Defaulting by Age Category

Metadata

Search Model

<All>

Measures

- Fact Application Details
- perc\_of\_defaulters

KPIs

- Client

Age Catg

- Dim Date
- Year

Dimension	Hierarchy	Operator	Filter Expression	Param...
<Select dimension>				

Year	Age Catg	perc_of_defaulters
2010	31-40	9.2
2010	41-50	7.4
2010	51-60	6.1
2010	above 60	4.5
2011	21-30	11.6
2011	31-40	9.4
2011	41-50	7.7
2011	51-60	6.6
2011	above 60	5.1
2012	21-30	11.7
2012	31-40	9.8
2012	41-50	8.1
2012	51-60	5.7
2012	above 60	5
2013	21-30	11.3

Solution Explorer

Solution Explorer (Ctrl+)

Search Solution Explorer (Ctrl+)

Solution 'Grad\_project' (1 project)

- Grad\_project
  - Data Sources
    - Loan Defaulter DWH.ds
  - Data Source Views
    - Loan Defaulter DWH.dsv
  - Cubes
    - Loan Defaulter DWH.cube
  - Dimensions
    - Client.dim
    - Loans.dim
    - Dim Date.dim
    - Previous.dim
    - Fact Application Details.dim
  - Mining Structures
  - Roles
  - Assemblies

Solution Explorer Team Explorer

Getting Started (SSIS)

# #Defaulters by age category perspective

The screenshot displays the Microsoft Visual Studio interface for an SSIS project named 'Grad\_project'. The main window shows the 'Loan Defaulter DWH.cube [Design]' view. The 'SSIS Toolbox' on the left contains a 'Defaulters by Age Category' cube and a 'Metadata' section with a search bar and a list of measures, KPIs, and dimensions. The 'Solution Explorer' on the right shows the project structure, including 'Data Sources', 'Data Source Views', 'Cubes', 'Dimensions', 'Mining Structures', 'Roles', and 'Assemblies'. The 'Error List' at the bottom shows an error: 'An error has occurred while loading Getting Started'.

**Defaulters by Age Category**

Year	Age Catg	TARGET
2010	21-30	569
2010	31-40	760
2010	41-50	559
2010	51-60	408
2010	above 60	146
2011	21-30	552
2011	31-40	763
2011	41-50	587
2011	51-60	451
2011	above 60	167
2012	21-30	572
2012	31-40	817
2012	41-50	605
2012	51-60	385
2012	above 60	159


























# Avg credit by income type perspective

The screenshot displays the Microsoft Visual Studio interface for a project named 'Grad\_project'. The main window shows the 'Loan Defaulter DWH.cube [Design]' view. The 'SSIS Toolbox' on the left contains a 'Metadata' section with a search bar and a list of items including 'AVG Credit By Income Type', 'Measures', 'Fact Application Details', 'Credit', 'KPIs', 'Client', 'NAME INCOME TYPE', and 'Dim Date'. The 'Solution Explorer' on the right shows the project structure, including 'Data Sources', 'Data Source Views', 'Cubes', 'Dimensions', and 'Roles'. The central data table, titled 'Avg Credit by Income Type', displays the following data:

Year	NAME INCOME TYPE	Credit
2010	"Commercial assoc...	678103.365498689
2010	"Pensioner"	547264.846698113
2010	"State servant"	676224.131401617
2010	"Student"	549159
2010	"Unemployed"	270000
2010	"Working"	580497.517459708
2011	"Businessman"	2250000
2011	"Commercial assoc...	668013.436351986
2011	"Pensioner"	533990.496763754
2011	"State servant"	667550.934311838
2011	"Student"	828695.25
2011	"Unemployed"	886500
2011	"Working"	573736.858595078
2012	"Businessman"	675000
2012	"Commercial assoc...	679641.081971831

# Risk KPI

A	B	C	D
Row Labels	TARGET	percentage	Risk Status
<b>2010</b>			
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	
<b>2011</b>			
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	
<b>2012</b>			
21-30	572	11.7	
31-40	817	9.8	
41-50	605	8.1	
51-60	385	5.7	
above 60	159	5	
<b>2013</b>			
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	
<b>2014</b>			
21-30	559	11.5	
31-40	805	9.6	
41-50	581	7.6	
51-60	406	6	
above 60	178	5.5	



Phase four

04

# Data visualization with interactive dashboards using Power BI:

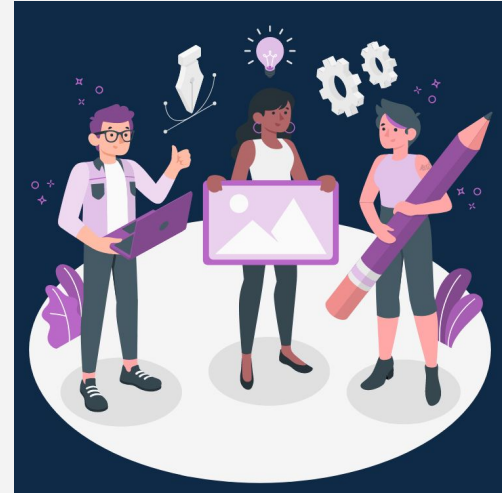


## Loan information insights

- 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.

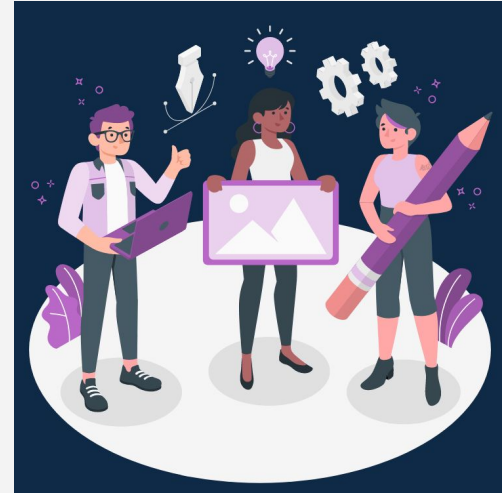
## Defaulters personal information

- 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:



A low-angle, upward-looking photograph of several modern skyscrapers with glass and steel facades. The buildings are set against a clear blue sky with some light, wispy clouds. A thin white rectangular frame is superimposed over the center of the image, enclosing the text.

Phase six

06

# Publishing the project on the Power BI service:

The screenshot displays the Microsoft Power BI service interface. At the top, the header bar shows 'Power BI My workspace' on the left, 'Grad\_proj | Data updated 1/27/22' in the center, and a 'Trial: 40 days left' notification on the right. Below the header, the left navigation pane lists various options: Home, Favorites, Recent, Create, Datasets, Goals, Apps, Shared with me, Deployment pipelines, and Learn. The main content area is titled 'Loan Defaulters Report' and features a background image of a calculator and coins. The report includes six interactive tiles: 'Loan Information' (with a hand holding a coin), 'Defaulters Personal Information' (with a person icon), 'Professional Information' (with a person icon), 'Financial Status' (with a wallet icon), 'Previous Loans Information' (with a left arrow icon), and 'Risks Information' (with a warning triangle icon). On the right side of the report, there is a 'Filters' section with a search bar and a message stating 'There aren't any filters to display.'

Power BI My workspace

Grad\_proj | Data updated 1/27/22

Trial: 40 days left

Search

File Export Share Chat in Teams Get insights

Pages

- Home
- Loan info
- Personal Info
- Professional Info
- Financial Status
- Previous
- Risks

## Loan Defaulters Report

- Loan Information
- Defaulters Personal Information
- Professional Information
- Financial Status
- Previous Loans Information
- Risks Information

Filters

Search

There aren't any filters to display.

**Thank you**