```
# STEP 1: DATA SETUP

# 1. Install and import libraries
import pandas as pd
import numpy as np


# 2. Upload files from your system
from google.colab import files
uploaded = files.upload()
# Upload:
# - Bank Customer Churn Prediction.xlsx
```

Choose Files Bank Custo…ediction.xlsx

- **Bank Customer Churn Prediction.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 678563 bytes, last modified: 4/5/2025 - 100% done

```
# 2. Upload files from your system
from google.colab import files
uploaded = files.upload()
# Upload:
# - loan details.xlsx
```

Choose Files loan details.xlsx

- **loan details.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 49107 bytes, last modified: 4/5/2025 - 100% done

```
# 3. Load the churn and loan datasets
churn_df = pd.read_excel("Bank Customer Churn Prediction.xlsx")
loan_df = pd.read_excel("loan details.xlsx")


# 4. Clean column names
churn_df.columns = churn_df.columns.str.strip().str.lower().str.replace(" ", "_")
loan_df.columns = loan_df.columns.str.strip().str.lower().str.replace(" ", "_")


# 5. Randomly select 582 customers from churn dataset
sampled_customers = churn_df.sample(n=len(loan_df), replace=False).copy()


# 6. Assign customer_id to loan_df
loan_df['customer_id'] = sampled_customers['customer_id'].values


# 7. Merge both datasets on customer_id
combined_df = pd.merge(loan_df, sampled_customers, on='customer_id', how='inner')


# 8. Save as Excel file
excel_filename = "bank_churn_loan_combined.xlsx"
combined_df.to_excel(excel_filename, index=False)

# 9. Download the Excel file
files.download(excel_filename)

# 10. Preview output
print("Combined dataset shape:", combined_df.shape)
combined_df.head()
```

25)

d	dependents	education	self_employed	applicantincome	coapplicantincome	loanamount	loan_amount_term	.
o	0	Graduate	No	5849	0.0	NaN	360.0	
s	1	Graduate	No	4583	1508.0	128.0	360.0	
s	0	Graduate	Yes	3000	0.0	66.0	360.0	
s	0	Not Graduate	No	2583	2358.0	120.0	360.0	
o	0	Graduate	No	6000	0.0	141.0	360.0	