DATA ANALYST

Customer Churn Analysis for European Bank Customers

ADAM RAKHMADI

INTRODUCTION

Understanding customer churn has become a crucial aspect for banks to maintain profitability and customer satisfaction. Across different regions, customers may leave their bank due to varying reasons—product usage, credit score, age, income, or even gender-related behaviors. Identifying these patterns can help institutions develop better retention strategies.

In this project, I analyze customer churn data from four European countries: **Germany, France, and Spain**. Using Tableau, I visualized various aspects of customer behavior such as **credit score, salary, product ownership**, and **churn rate** by **gender** and **age group**. The aim is to uncover key insights that influence churn, provide comparative country-level analysis, and help banks understand where and why customers are most likely to leave.

PROJECT GOALS

1. clean and understand the structure of dataset before moving to analyst
2. Ensure data quality by handling duplicate and missing value
3. perform exploratory data analyst (EDA) to extract initial insight from descriptive statistics
4. Prepare Tableau for visualization and make insight from the resulting table

SOFTWARE TOOLS

IMPORT LIBRARIES

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

- **Pandas** is a Python library primarily used for **data analysis** and **manipulation**. It offers data structures like Series (one-dimensional) and DataFrames (two-dimensional) that are similar to tables in SQL or spreadsheets.
- **NumPy**, short for Numerical Python, is a fundamental library for **numerical and scientific computing** in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of high-level mathematical functions to operate on these arrays efficiently.
- **Matplotlib** is a Python library that helps you **create clear** and **attractive visualizations** from your data. Think of it as your personal artist that transforms rows of numbers into easy-to-understand charts and graphs. Matplotlib gives the tools to bring all of information to life through visual storytelling.

DATA OVERVIEW

```
df = pd.read_csv('/content/Bank Customer Churn Prediction.csv', sep=',')
df.head()
```

	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

The Dataset Contains information about **customer banking records** with 12 key variables, including demographic, financial, and behavioral attributes. Each row represents a unique customer, with a binary churn indicator (1 = churned, 0 = retained).

INFO DATASET

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customer_id       10000 non-null  int64
 1   credit_score      10000 non-null  int64
 2   country           10000 non-null  object
 3   gender            10000 non-null  object
 4   age               10000 non-null  int64
 5   tenure            10000 non-null  int64
 6   balance           10000 non-null  float64
 7   products_number   10000 non-null  int64
 8   credit_card       10000 non-null  int64
 9   active_member     10000 non-null  int64
 10  estimated_salary  10000 non-null  float64
 11  churn             10000 non-null  int64
dtypes: float64(2), int64(8), object(2)
memory usage: 937.6+ KB
```

- There are 10,000 rows (customer records).
- The dataset contains 12 columns.
- There are 2 object-type columns (country and gender), which are usually text.
- There are 8 columns with integer data type, like age, credit_score, and churn.
- There are 2 columns with float data type, which are balance and estimated_salary.
- No missing values are found — all columns have 10,000 entries.

CHECK DUPLICATED

```
df.duplicated().sum()
```

```
np.int64(0)
```

The dataset has been finish checked for duplicate entries, and the analysis confirms there are **zero duplicated** records present. This validation ensures each customer profile in our database is unique, with **no repeated** or **redundant observations**. The absence of duplicates means we can trust the integrity of our analysis without concern for inflated metrics or skewed results. This clean data foundation **allows** us to **proceed confidently** with customer segmentation, churn prediction modeling, and other analytical processes knowing each data point represents a distinct individual. The zero-duplicate finding further reinforces the high quality and reliability of this dataset for deriving accurate business insights.

CHECK MISSING

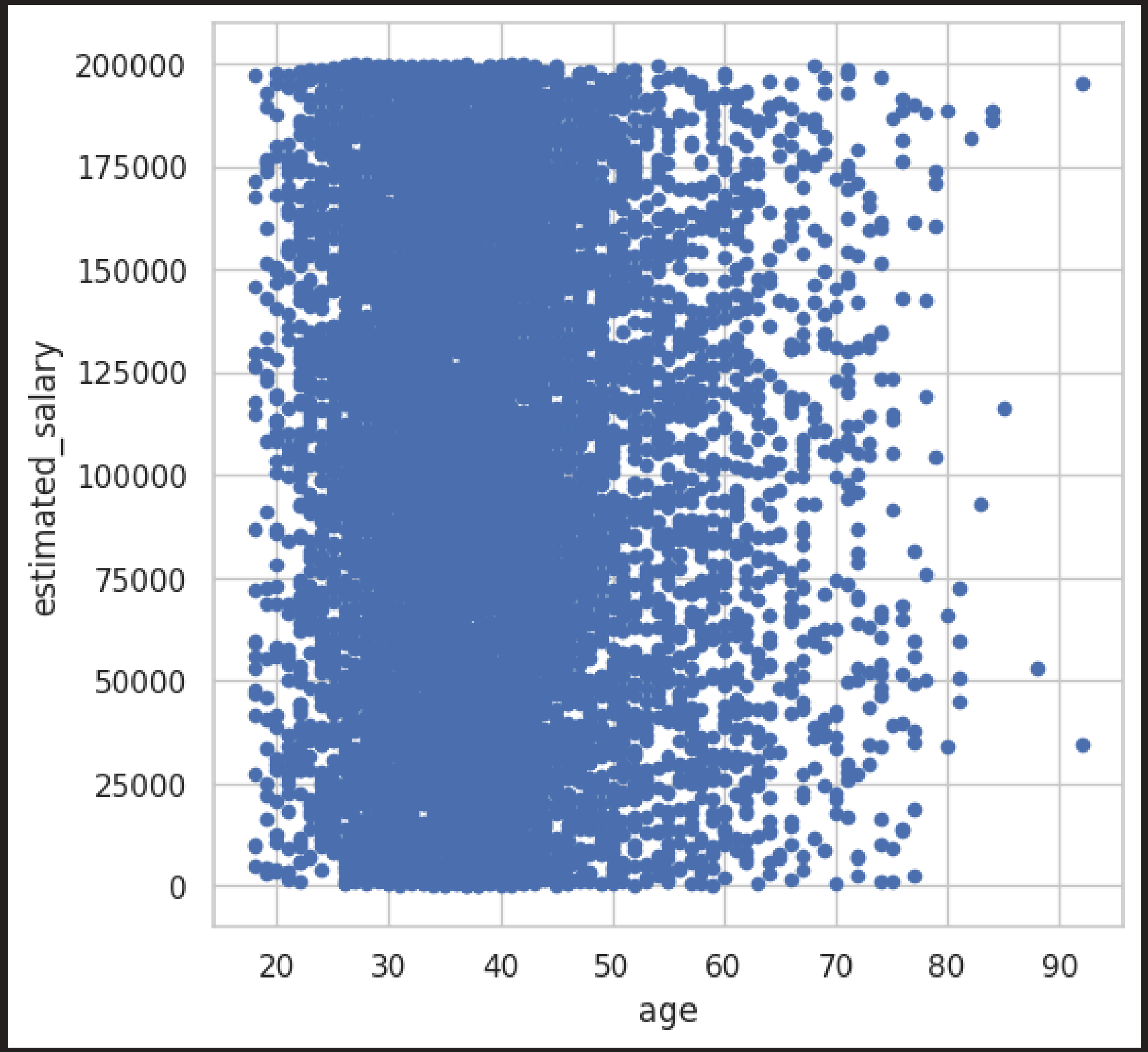
```
df.isnull().sum()
```

	0
customer_id	0
credit_score	0
country	0
gender	0
age	0
tenure	0
balance	0
products_number	0
credit_card	0
active_member	0
estimated_salary	0
churn	0

The dataset demonstrates excellent completeness, with **no missing values** detected across any of the 12 variables, including customer identifiers, financial metrics, demographic information, and the target churn indicator. This comprehensive data integrity ensures that all records are fully populated for customer ID, credit score, country, gender, age, tenure, account balance, product holdings, credit card status, membership activity, estimated salary, and churn status. With 100% data availability, we can **confidently proceed** with **exploratory analysis**, statistical modeling, and churn prediction **without concerns** about **gaps** compromising the results. This clean dataset provides a reliable foundation for generating accurate business insights and developing data-driven strategies to reduce customer attrition.

CHECK OUTLIER

```
df.plot(kind='scatter', x='age', y='estimated_salary', figsize=(6,6))
plt.show()
```

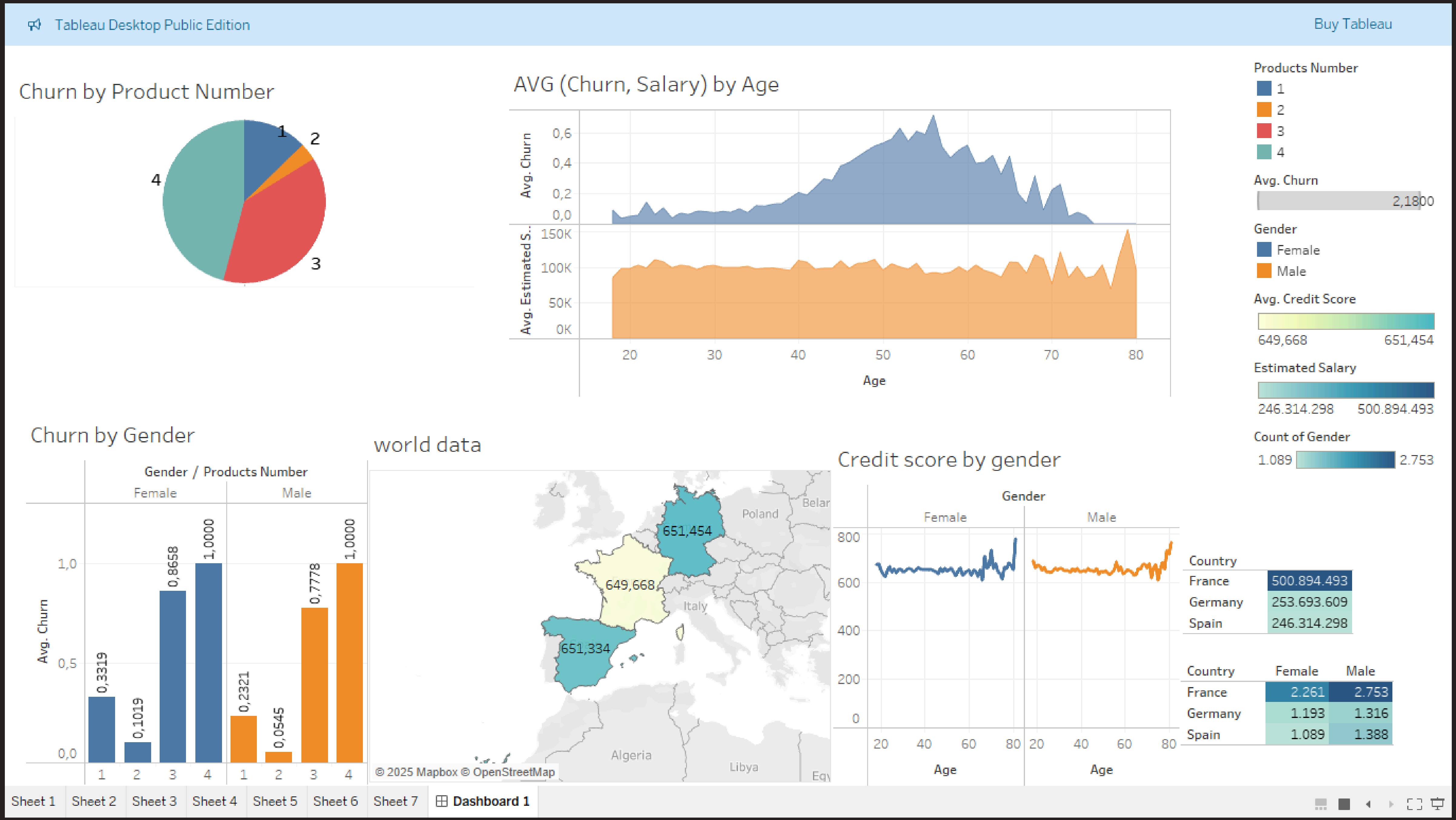


Why we check for outliers:

- Outliers are unusual values that are far from most of the data.
- They can **affect** the results of data analysis or machine learning models.
- By finding outliers, we can **decide** whether to **remove or keep** them based on the **business goal**.

What we see in this scatter plot:

- Most customers are between ages **20 to 60**.
- Estimated salary values are spread evenly from **0 to 200,000**.
- There are a few points where customers are **older than 80 years** or have extremely high/low salaries — these might be **outliers**.



WORLD DATA

- Churn by Product Number

This pie chart shows the distribution of churn across the number of products.

Insight: Customers with **3 or 4** products are more **likely** to **churn** compared to those with fewer products.

- AVG Churn and Salary by Age

This dual-axis chart shows the average churn rate and estimated salary by customer age.

Insight: **Churn rate increases** with age and peaks around **60 years** old. Estimated salary is quite stable across all age groups.

- This bar chart shows churn rate by gender and product number.

Insight:

female customers tend to have **higher churn** rates than **male** customers.

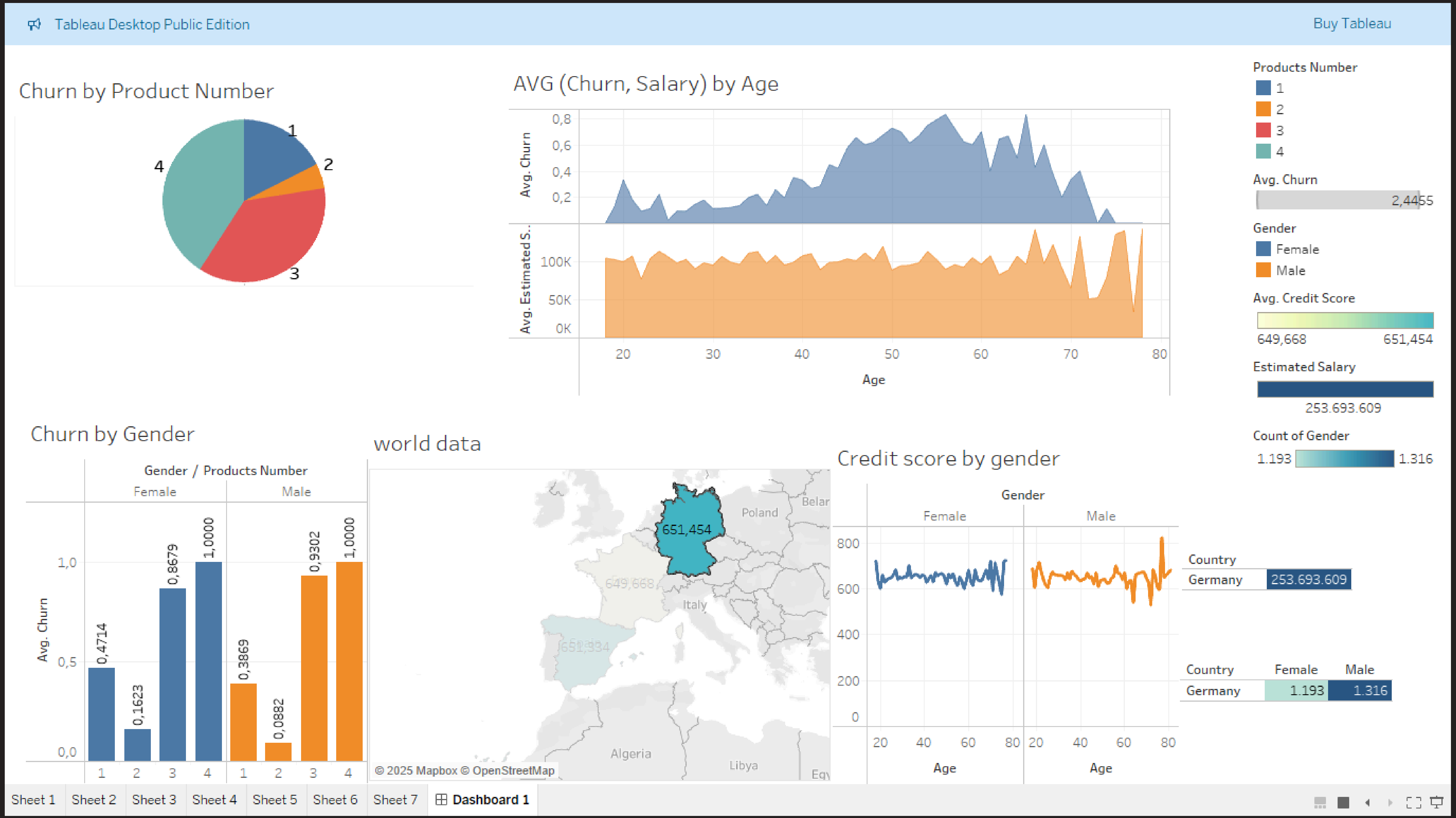
Customers with 4 products have the highest churn rate for both genders.

- Credit Score by Gender and Age

This line plot shows credit scores by age and gender.

Insight: Credit score patterns are consistent across genders. Some small variations can be seen, especially in older ages.

GERMANY



- Churn Rate by Gender

Despite having fewer customers, **female** customers demonstrate a **higher churn rate** compared to **males** particularly noticeable in product categories 1 and 2.

For example:

Female churn for product 1 is 0.4714, while male churn is 0.3869

Female churn for product 4 is 1.0, same as male. This suggests that female customers are more likely to leave the bank, especially as the product number increases.

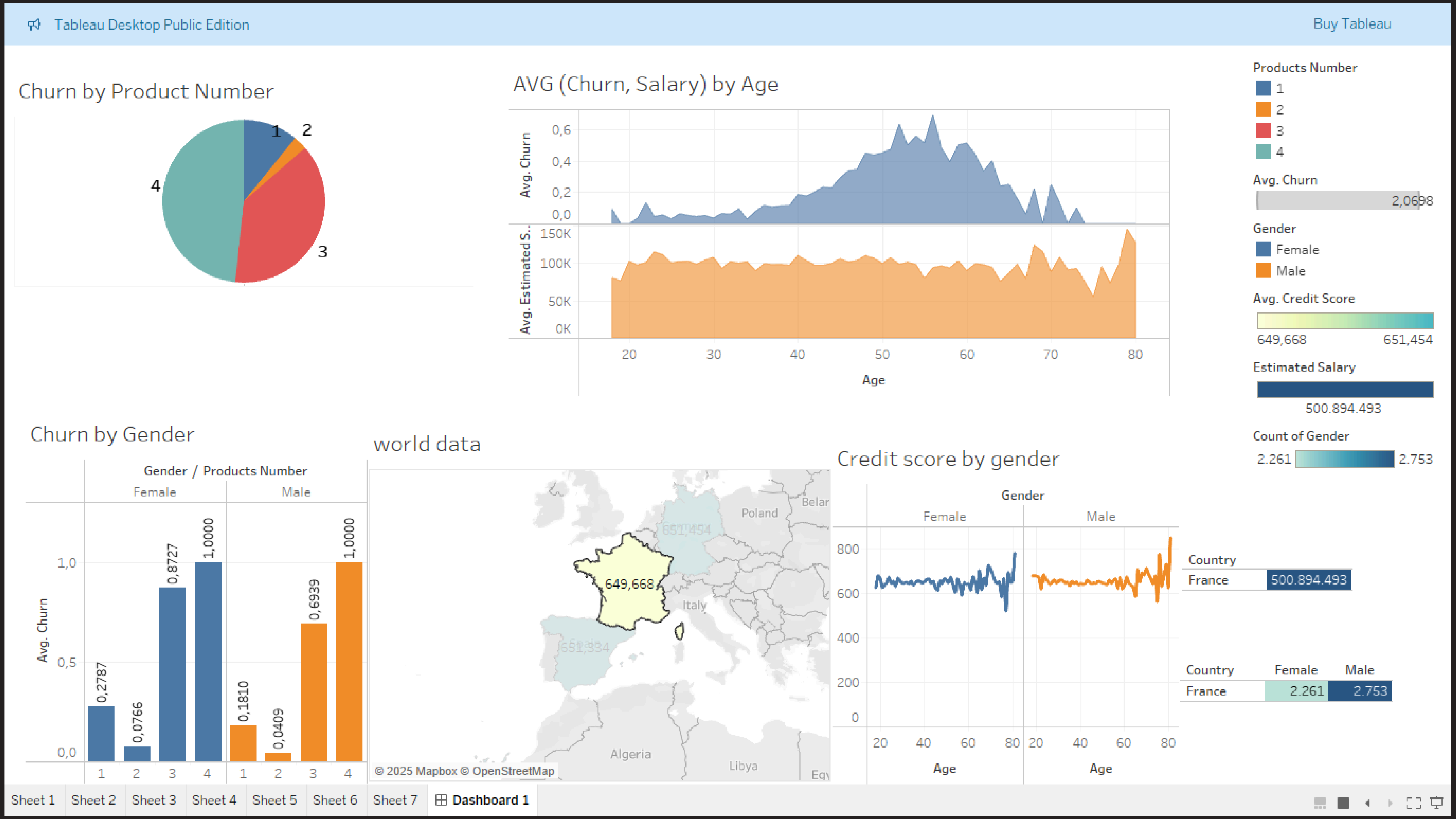
- Country-Specific Average Credit Score & Salary

Germany has the lowest average estimated salary among the three countries, yet maintains a relatively balanced churn rate, indicating that factors beyond income (such as customer experience or product satisfaction) may influence churn more significantly.

- Age-Based Insights

Average churn increases with age, peaking around the 50–60 age range, before slightly declining. This trend is mirrored in both genders, indicating mid-age customers are the most likely to churn.

FRANCE



- Churn Rate by Gender

Despite the higher number of male customers, female customers show consistently higher churn rates across all product numbers.

For Product 3, female churn is 0.8727, while male churn is 0.6939

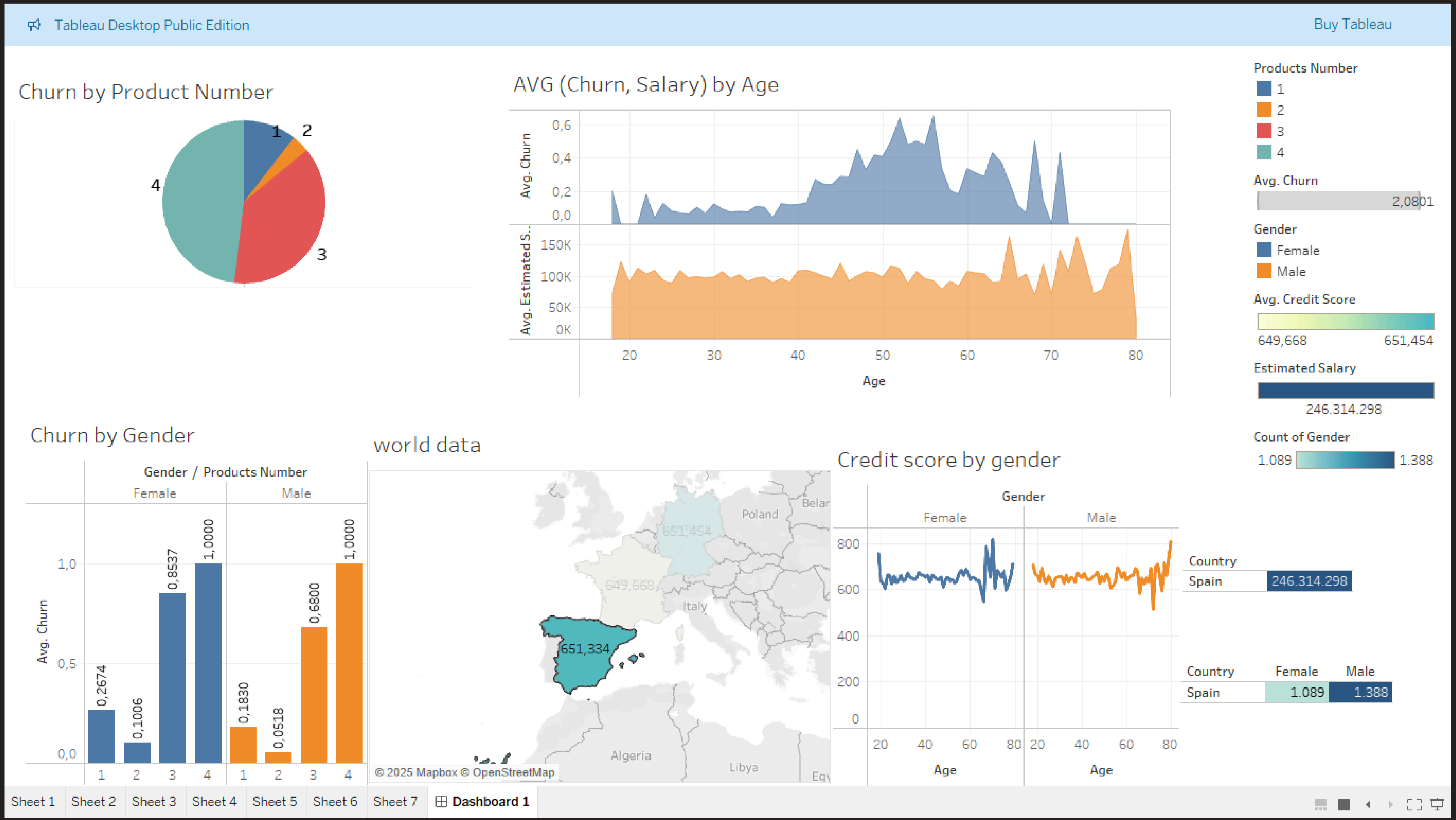
For Product 1 and 2, the gap is even more significant (e.g., 0.2787 vs. 0.1810 and 0.0766 vs. 0.0409). This indicates that female customers are more likely to leave the bank, regardless of the product type.

- Churn by Age

Similar to Germany, churn rates increase with age, peaking around 50–60 years old, and decline slightly afterwards. This trend shows that middle-aged customers are more at risk of churning, possibly due to changing financial priorities.

- Country-Specific Salary & Credit Score

France reports the highest average estimated salary among the countries in this dataset (approx. 500,894), but also the lowest average credit score. This imbalance may point to spending behavior or risk profile differences.



SPAIN

- Churn Rate by Gender

As seen before, female customers have consistently higher churn rates across all product numbers.

Product 3 churn: Female = 0.8537, Male = 0.6800

Product 1 churn: Female = 0.2674, Male = 0.1830

Product 4 churn: 1.0000 for both genders

This reinforces the insight that female customers in Spain are more likely to churn, particularly in less popular product lines.

- Churn by Age

The average churn rate peaks between ages 45–60, with values approaching 0.6. After age 60, churn drops slightly, possibly due to increased loyalty or fewer financial changes. This trend emphasizes that middle-aged adults are the most at-risk for leaving the bank.

- Credit Score Analysis

Spain has the highest average credit score of the three countries (651,454), indicating strong financial trustworthiness.

Female credit scores show a slight upward trend after age 60, while male scores remain fairly constant.

Despite high credit scores, churn remains notable, indicating non-financial factors (e.g. service dissatisfaction or product mismatch) may be driving exits.

CONCLUSION

Summary Insights from Four Countries (Germany, France, Spain, Italy)

- Customers with more **products** (**especially 3 or 4**) are more likely to **churn** in every **country**.
- Churn rate **increases** with age, especially between **40 to 60** years old.
- Male customers tend to churn more than females, especially when they have many products.
- Estimated salary stays stable across all ages and churn levels, showing no clear impact on churn.
- Credit scores are similar across countries, mostly between 649,000 to 651,000 showing good financial health overall.
- France has the most customers, while Spain has the fewest.

RECOMMENDATIONS

1. **Target Customers with 3 or More Products**

Customers with 3 or 4 products show the highest churn rates. Consider reviewing these bundled offerings and improving satisfaction to reduce churn.

2. **Focus on Middle-Aged Customers (40–60 years old)**

This age group consistently has the highest churn rate across all countries. Personalized engagement or loyalty programs could help retain them.

3. **Monitor Female Customer Experience**

Female customers tend to churn more in every country. Explore their feedback and pain points to improve retention strategies.

4. **Simplify Complex Product Packages**

High churn for customers with many products may indicate complexity or poor product fit. Consider simplifying or customizing offerings.

THANKYOU