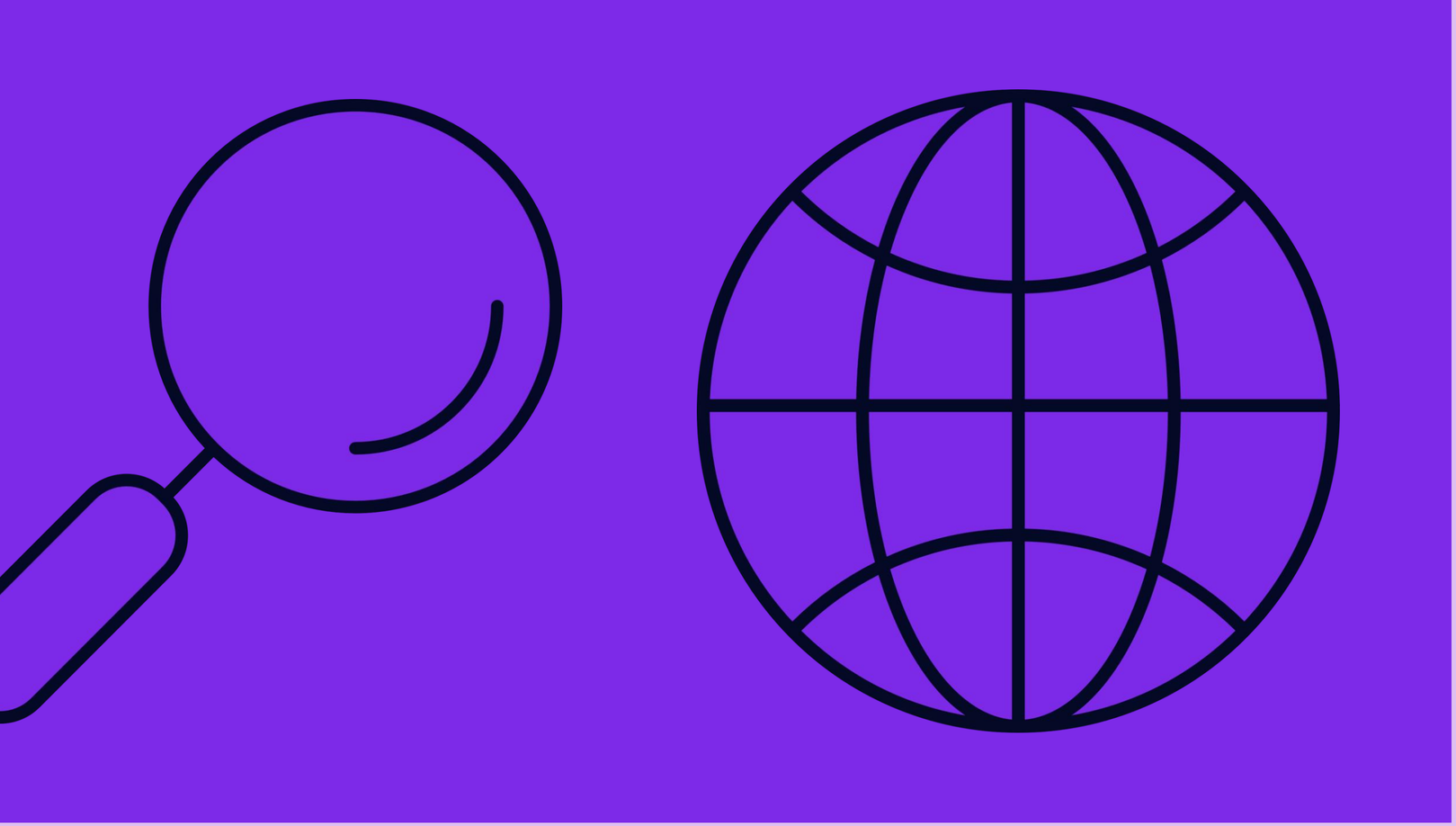
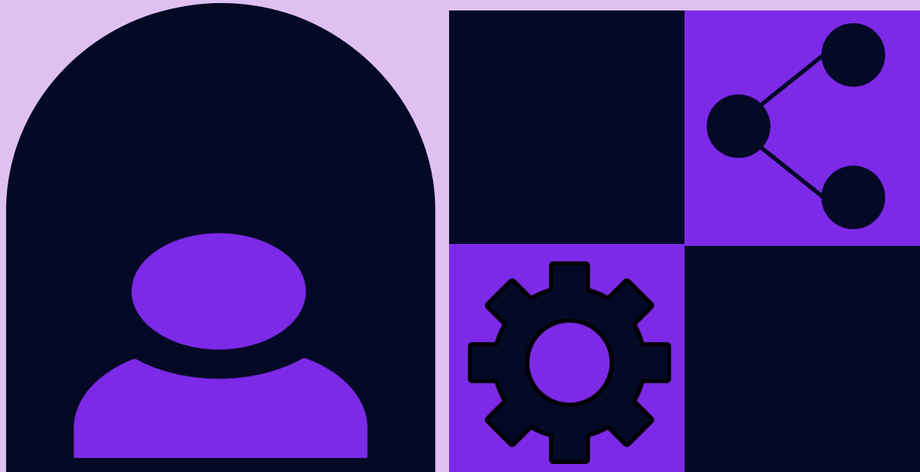


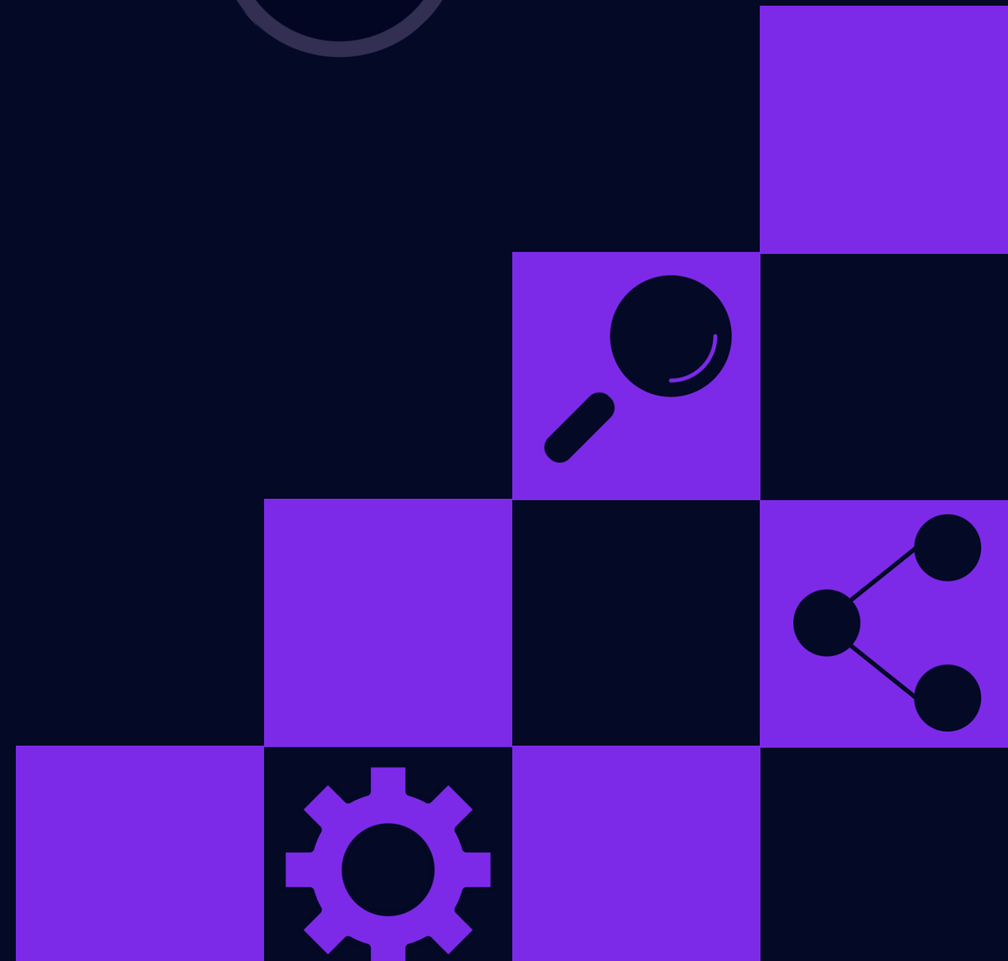
Team: jAIm (Jan Ritt, Obermüller Imre) **Date:** 10/01/2025

Spotify Analysis Dataset 2025



Data Science & AI

Search for data & basic analysis



Workflow

Find data

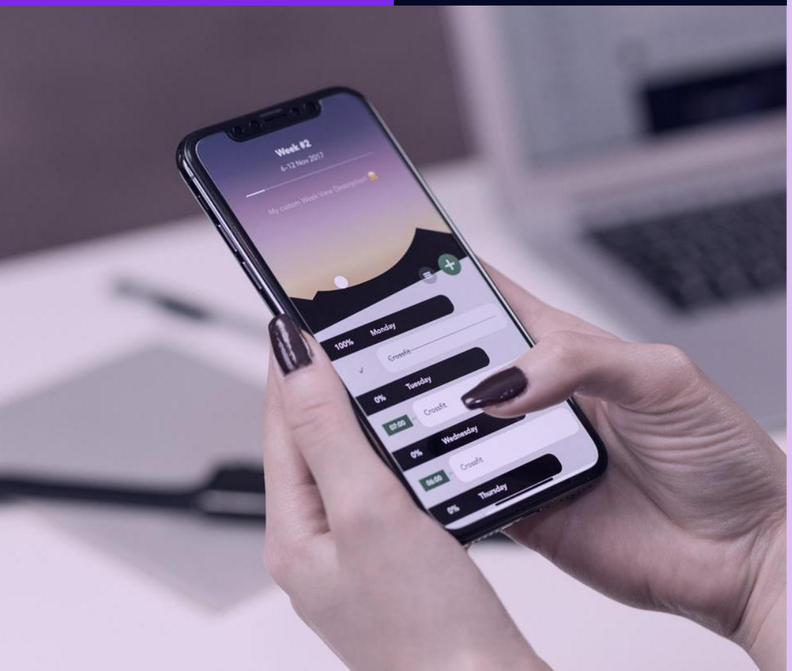
Find data from www.Kaggle.com or www.data.gv.at

Analyze metadata

Which metadata exists and what is the purpose of each?

Examine data structure

Which rows and columns exist? Classification-column?



Findings

Churn rate and class balance:

- About 26% churn (2,071 churned vs 5,929 active out of 8,000). The target is moderately imbalanced, so report metrics beyond accuracy..

Usage intensity vs churn::

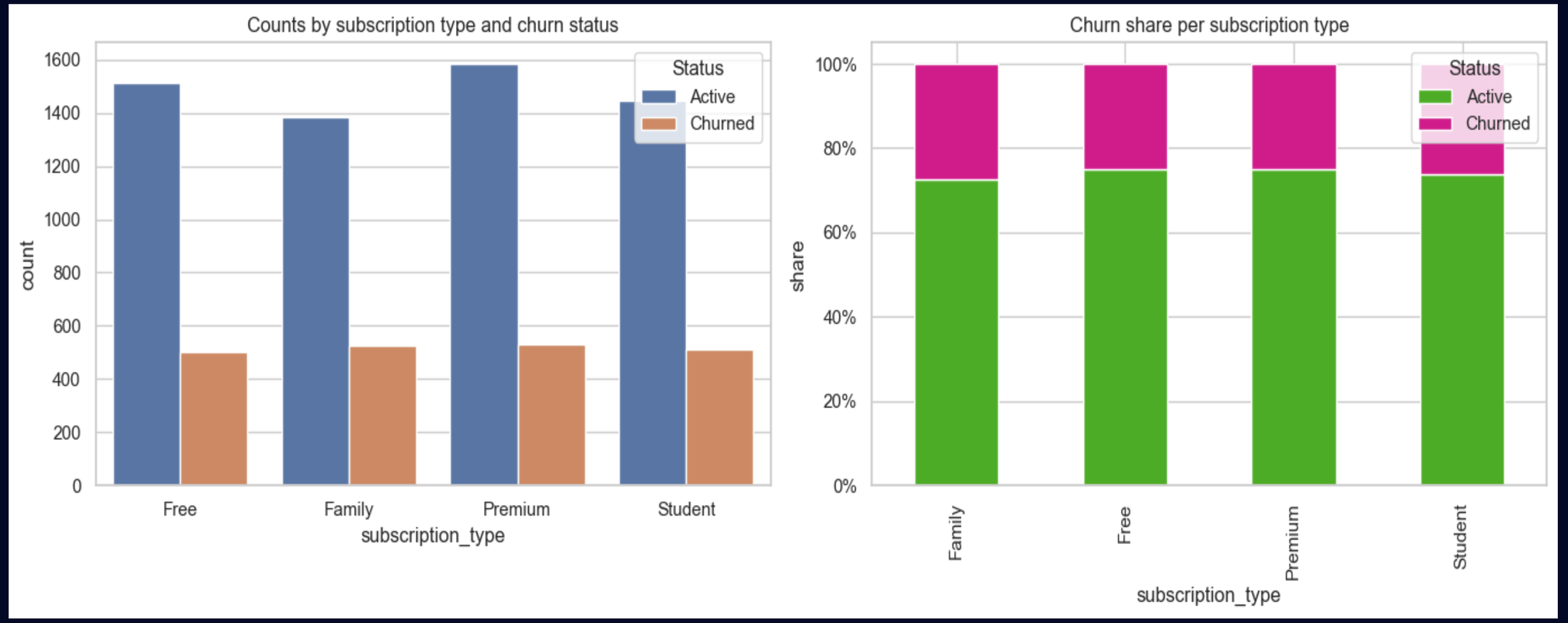
- Boxplots indicate churned users have lower typical usage (lower listening_time and songs_played_per_day) than active users, suggesting engagement is a strong signal for churn.

Subscription type matters:

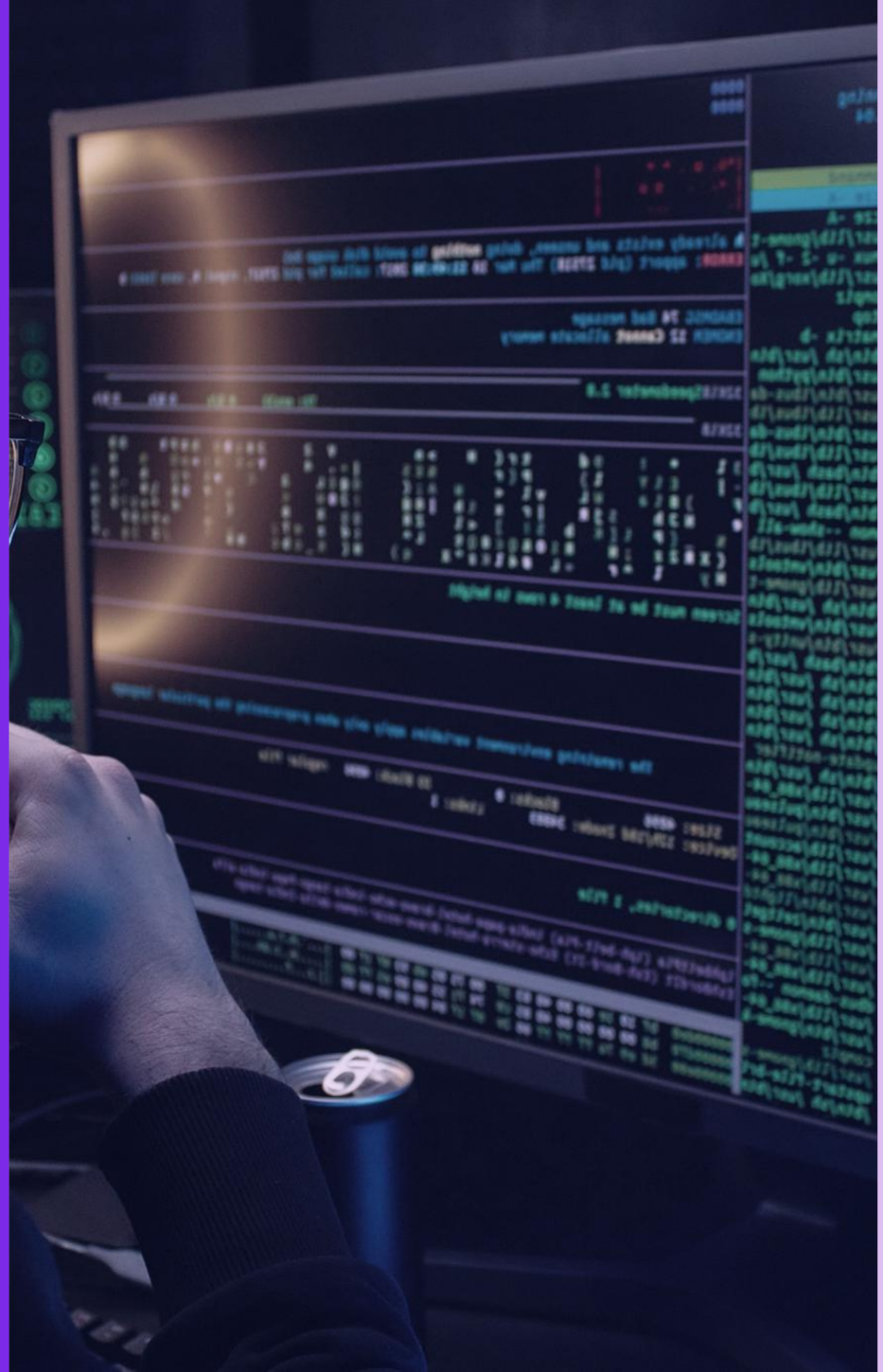
- The churn-by-subscription visuals show notable differences across plans; churn share is highest for lower-tier plans and lowest for higher-tier plans, indicating plan segmentation is predictive.

Active vs. Churned

Counts of active vs. churned per subscription_type



Summary

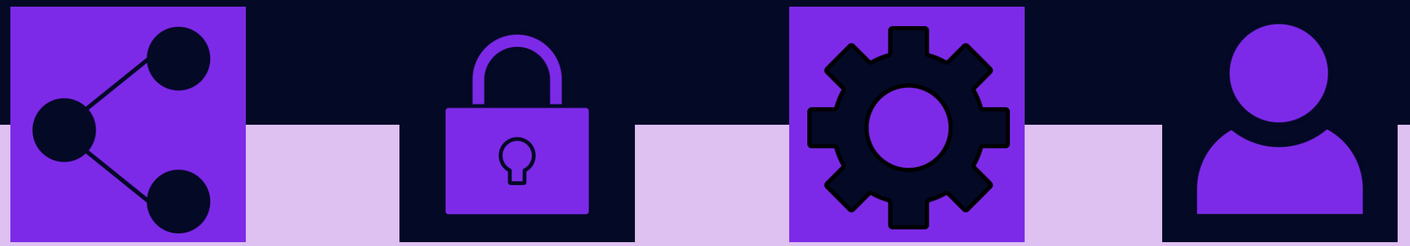


We validated the Kaggle Spotify churn dataset, created and executed an EDA notebook. Generated visuals: target distribution, engagement by churn, churn by subscription.

Key takeaways:

Churn is ~26%, churned users show lower engagement, and lower-tier plans have higher churn.

Thank you!



Any questions? Feel free to ask!

jAlm

HTL Leonding
next level