



# Austrian Startup Project

---

Lumos Data Science

by:  
Anton and Michi



# What is Lumos?

---

Lumos is a Vienna-based student data science consulting initiative.

- Our Team
  - 30+ students from different studies
- Our Expertise
  - Software Engineering
  - Data Science
  - Economics
  - Statistics...
- What we deliver
  - Machine Learning
  - Data Analytics
  - Business Intelligence

# Project Overview



Advise a European VC firm on understanding the Austrian startup market

## Client Context

- European VC firm exploring investment opportunities in Austria.
- Lacks knowledge and data-driven insights.

## Our Role

- Act as student data science consultants
- Provide strategic guidance based on data and predictive modeling

## Our Approach

- Scraped and enriched data from Austrian startup listings
- Created features to analyze funding status and activity
- Built a predictive model to identify factors linked to startup funding

# Our Deliverables

---



## Custom Dataset

- 500+ Austrian startups
- Enriched with variables using **OpenAI + SerperDev** APIs

## Market Analysis

- Visual and statistical insights
- Overview of the Austrian Startup Market

## Predictive Model

- **Random Forest** classification
- Highlights key factors driving funding success

# The dataset

---



## 500+ Austrian Startups

- **Basic Info:** name, website, location...
- **Social Media:** LinkedIn, Instagram
- **Company Details:** age, funding
- **Headlines:** mentions of startup in news
- **Description-Based:** quality, jargon...
- **Target Variable:** status of startup
  - inactive
  - no\_funding (but active)
  - funding

# Extraction Part 1: Scraping



## Scraping & Cleaning & Aggregation:

Home > prop.ID – simply real estate

### prop.ID – simply real estate



**Category:** Austria

**Business Description:**

prop.ID digitalizes standard processes for housing companies: Less work and time expenditure through hybrid or digital home owners meetings. prop.ID - simply real estate!

**Based in:** Vienna

**Tags:** real estate, digitalisation, owner meetings, property management

**Total Funding:** No funding announced yet

**Founded:** 2022

**Website:** <https://www.prop.id/>

# Extraction Part 2: Enriching the dataset

---



## Chat GPT API:

- Extracted text-based variables from descriptions
  - e.g. writing score, founder signal, market readiness
- Processed news headlines
  - e.g. financing, bankruptcy...



## Seper Dev API:

= Google Seach API

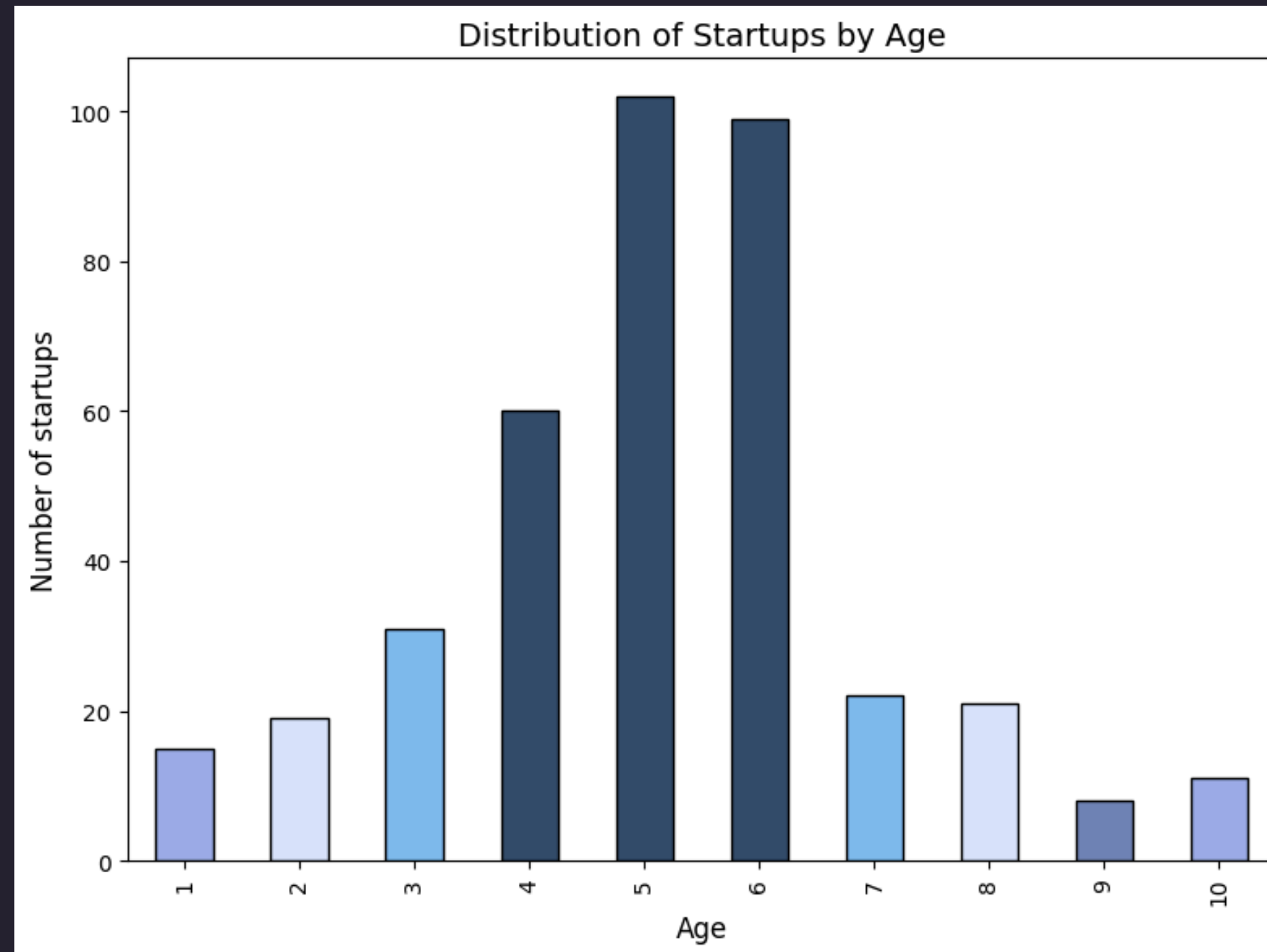
- Found social media handles via smart search (site:instagram.com)
  - e.g. got social media handles
- Searched for news articles and analyzed top 10 headlines:
  - e.g. Count of mentions
  - Funding/acquisition/bankruptcy indicators



# Data analysis



Main dataframe: web-scraped data on 680 Austrian startups  
after processing  $\Rightarrow$  388 startups of various ages across 9 regions in 15 categories

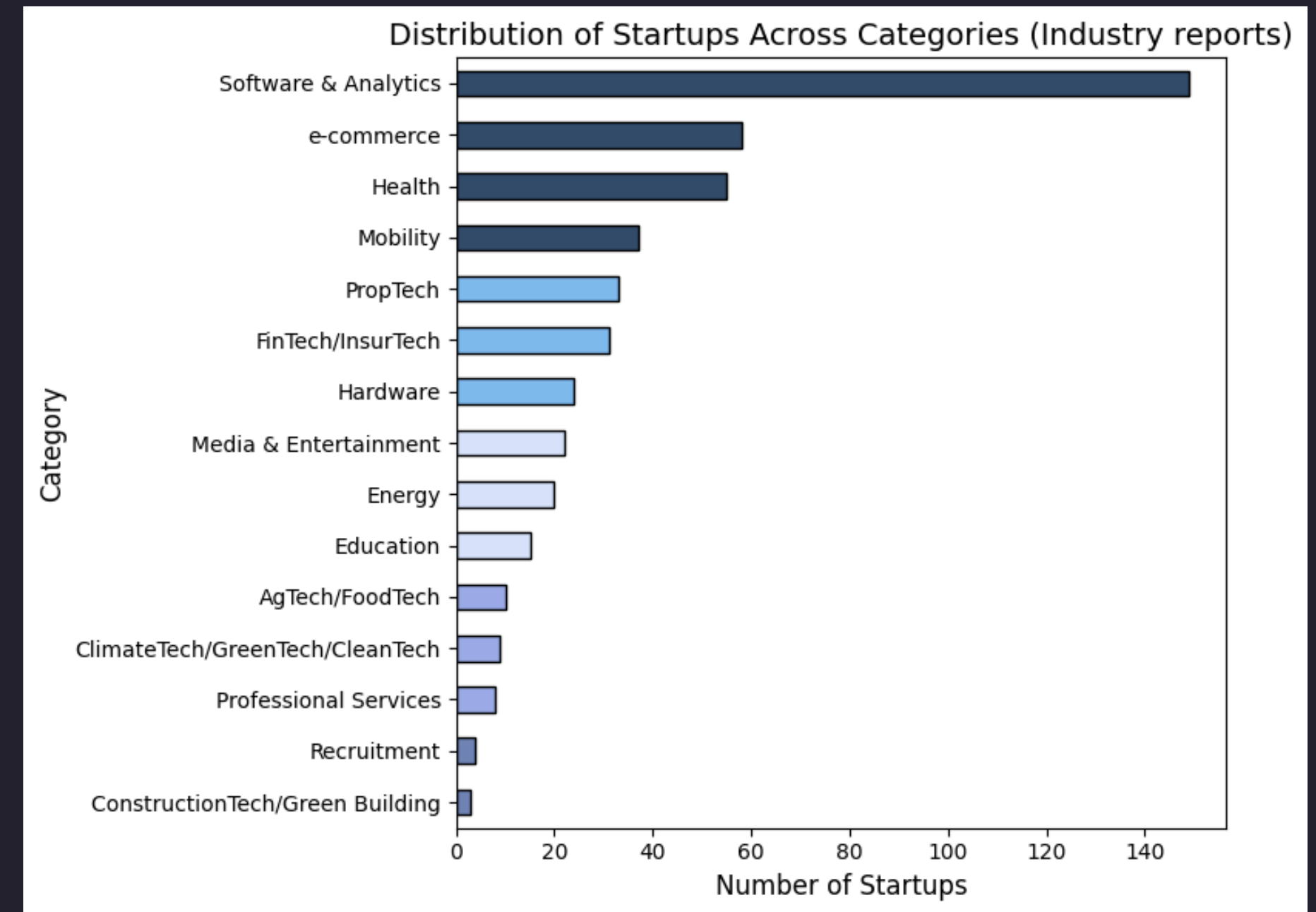
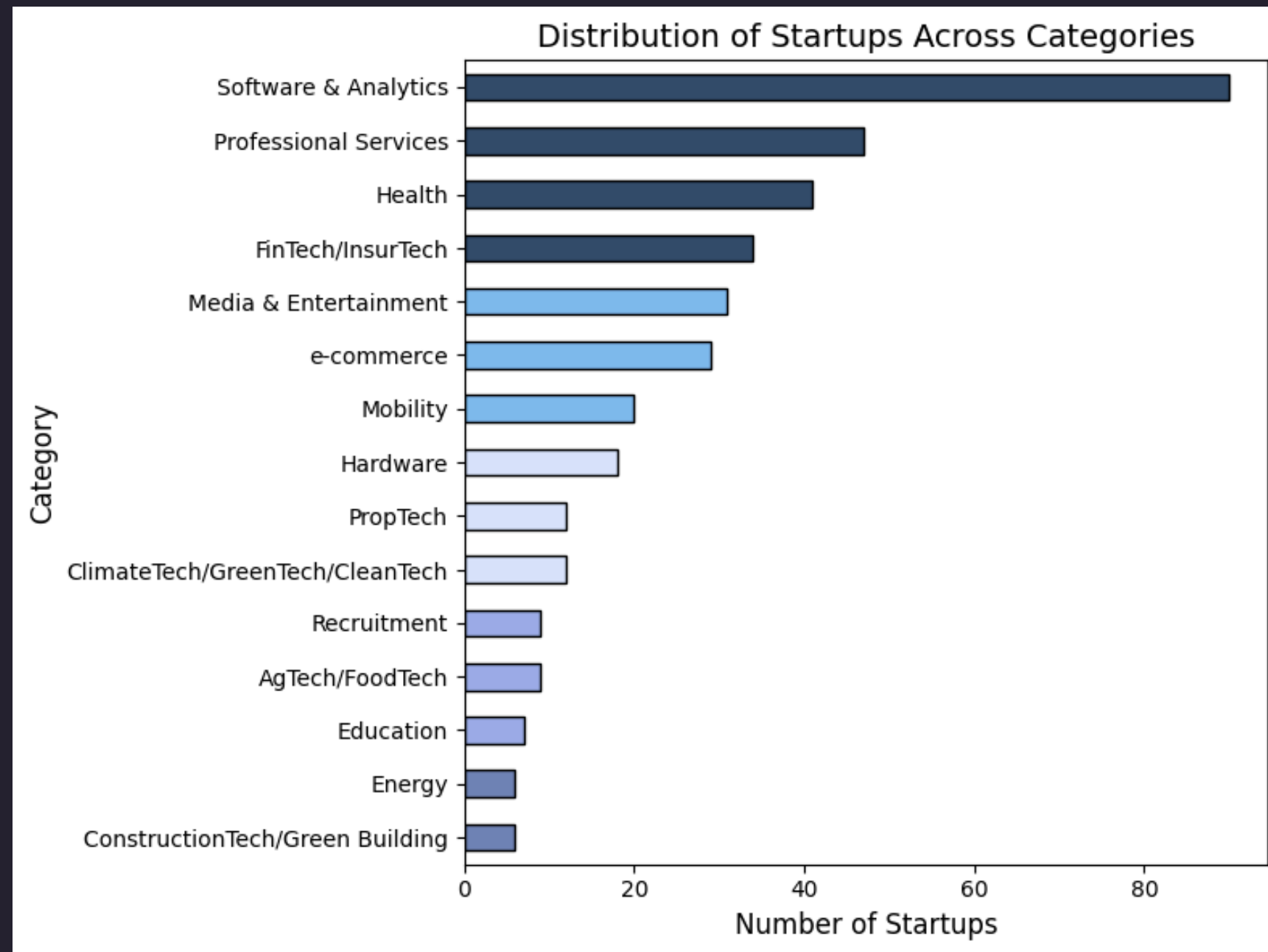




# Data analysis



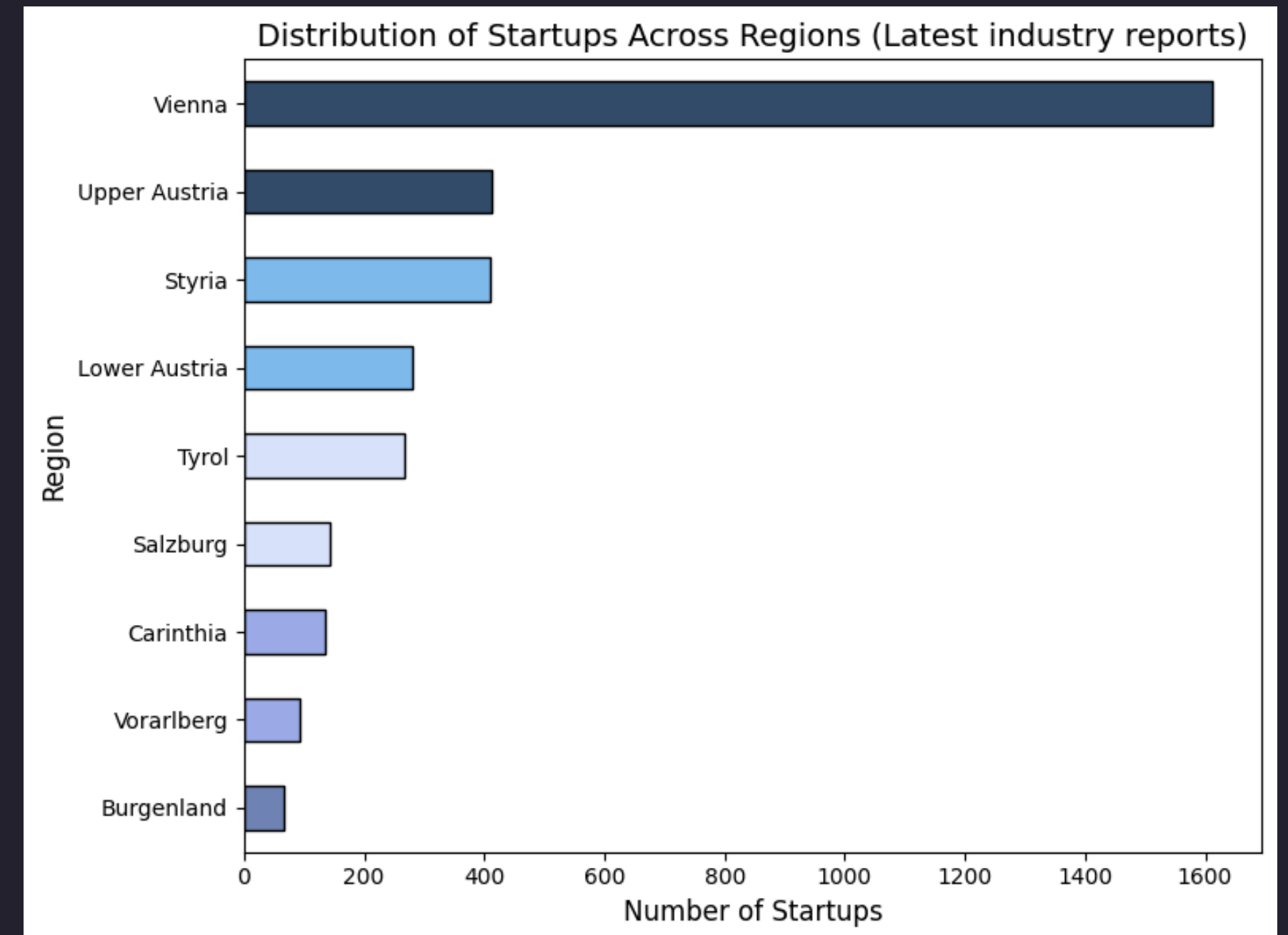
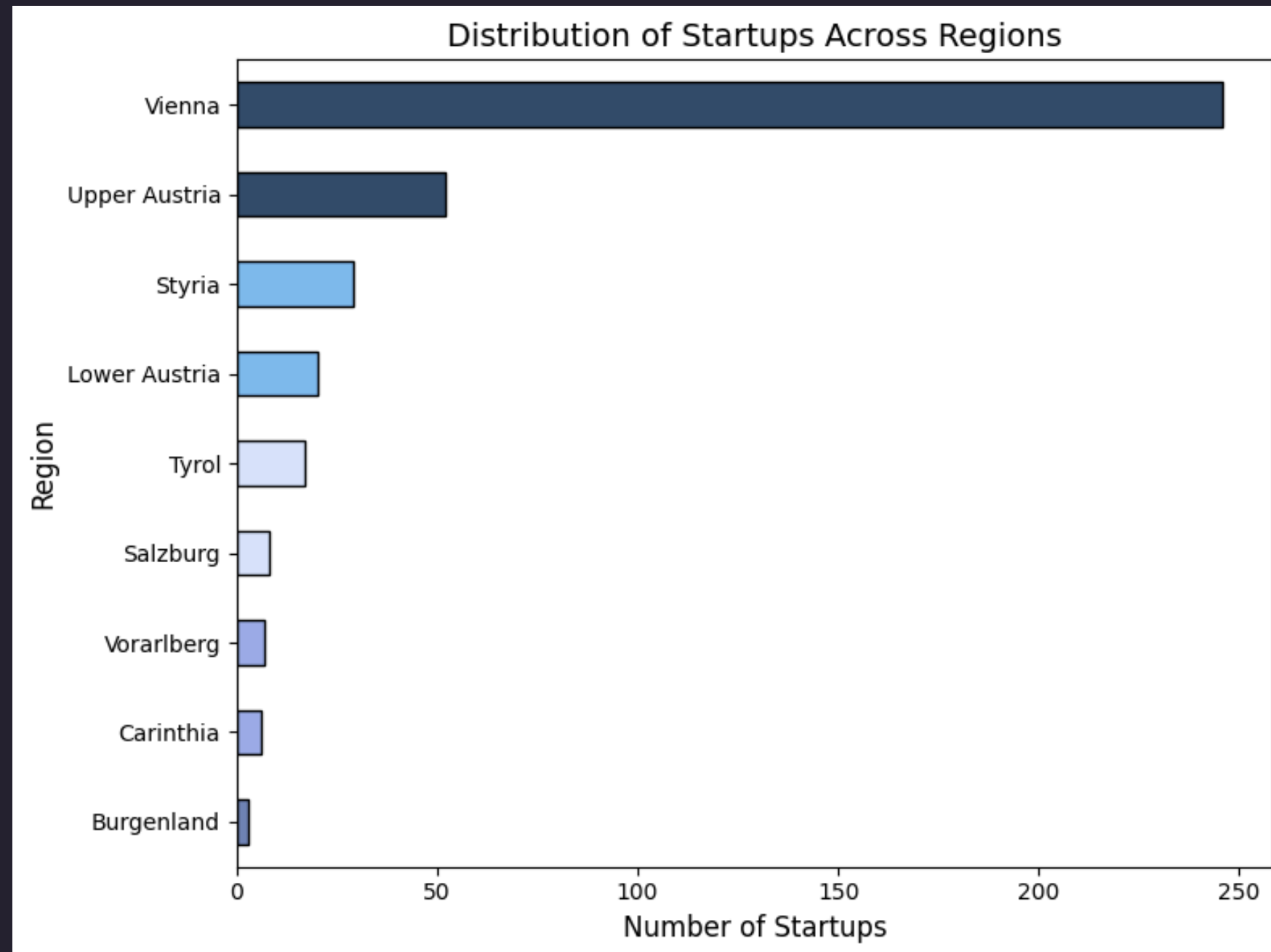
To the main dataset we added data from industry-wide reports and economic variables to see if regional socio-economic measures affect startup industry development



# Data analysis



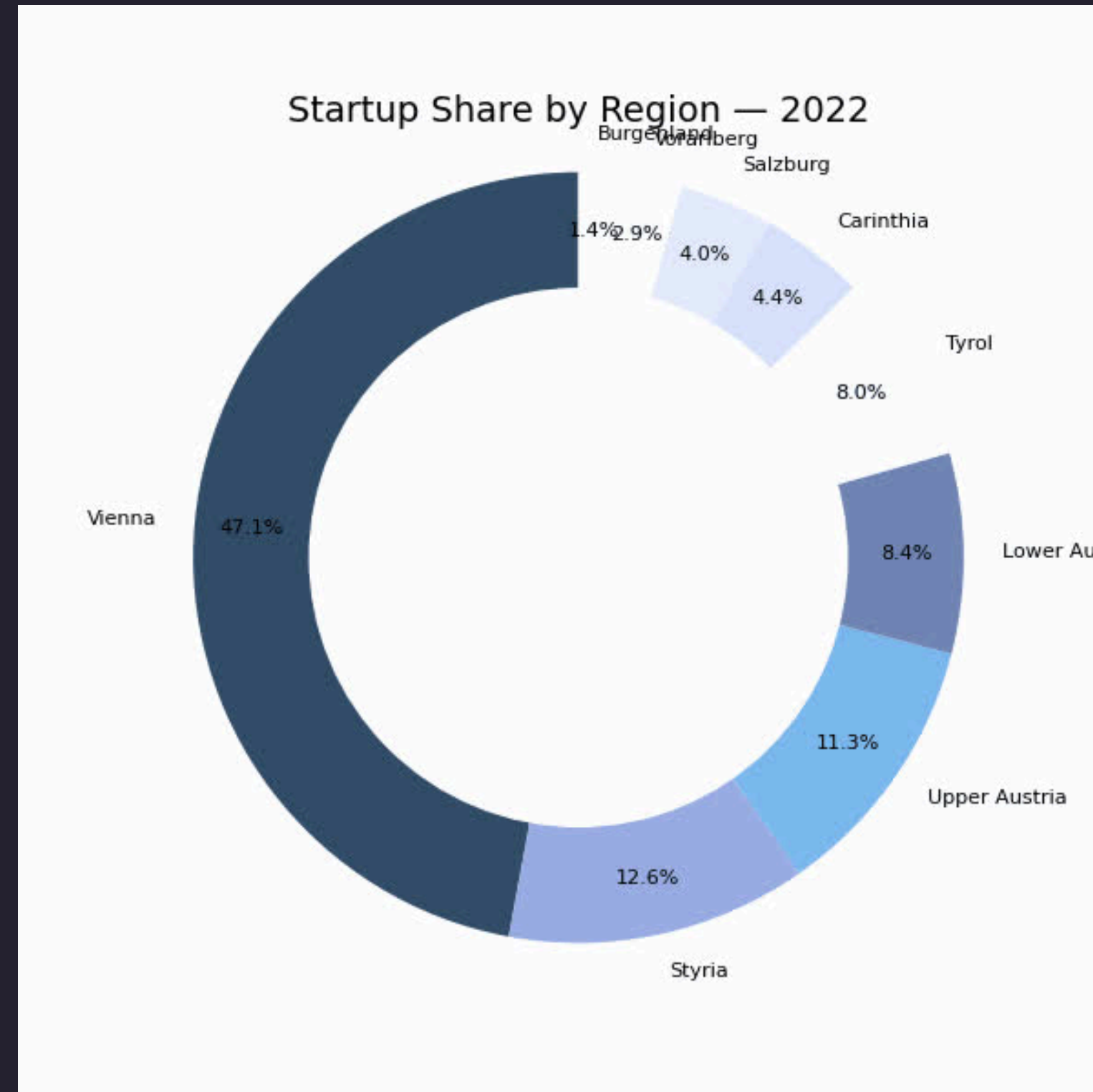
To the main dataset we added data from industry-wide reports and economic variables to see if regional socio-economic measures affect startup industry development



# Data analysis



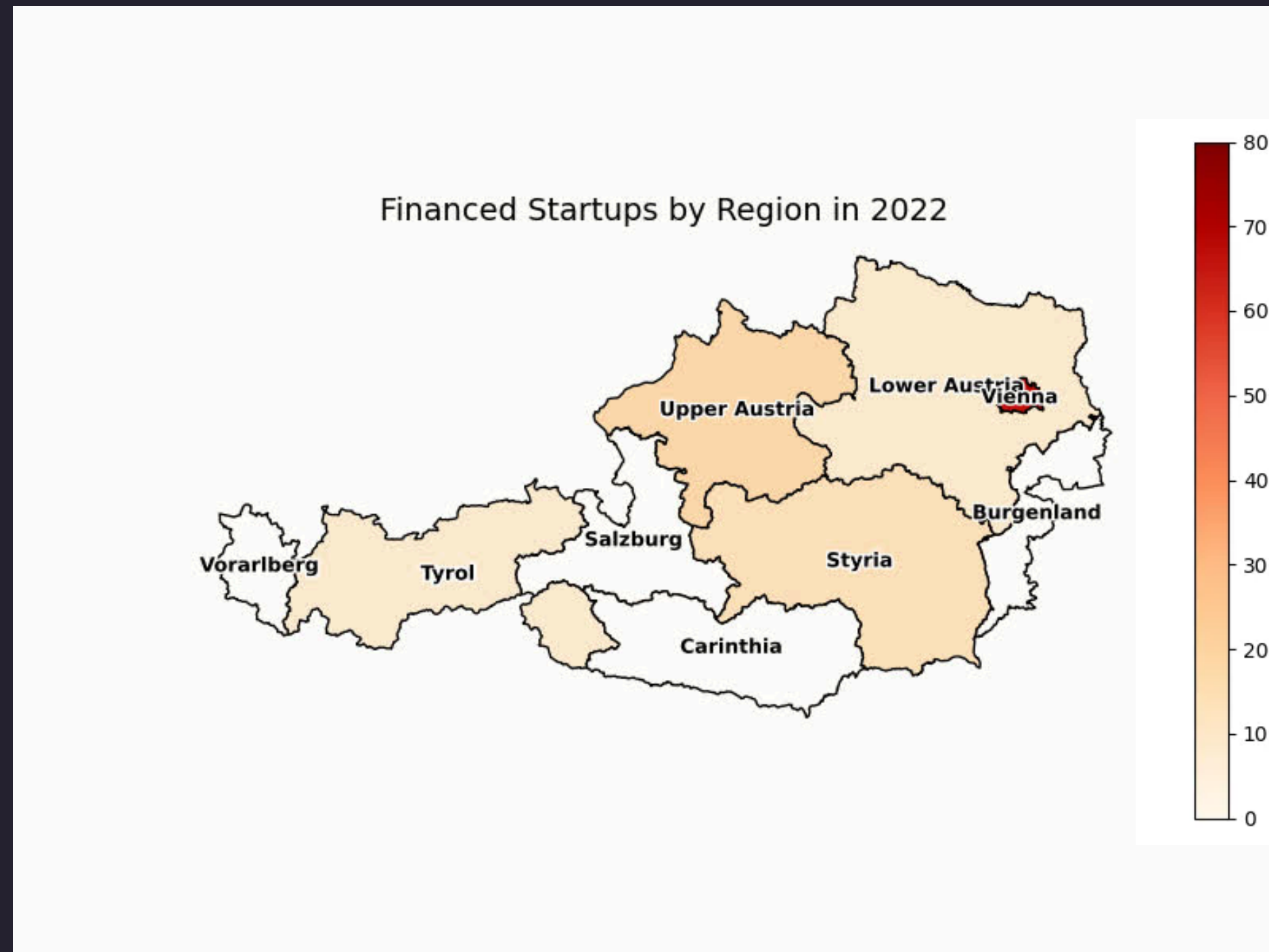
Vienna's share in total number of Austrian startups decreases, while regions like Styria and Upper Austria are closely competing for 2nd and 3rd places



# Data analysis



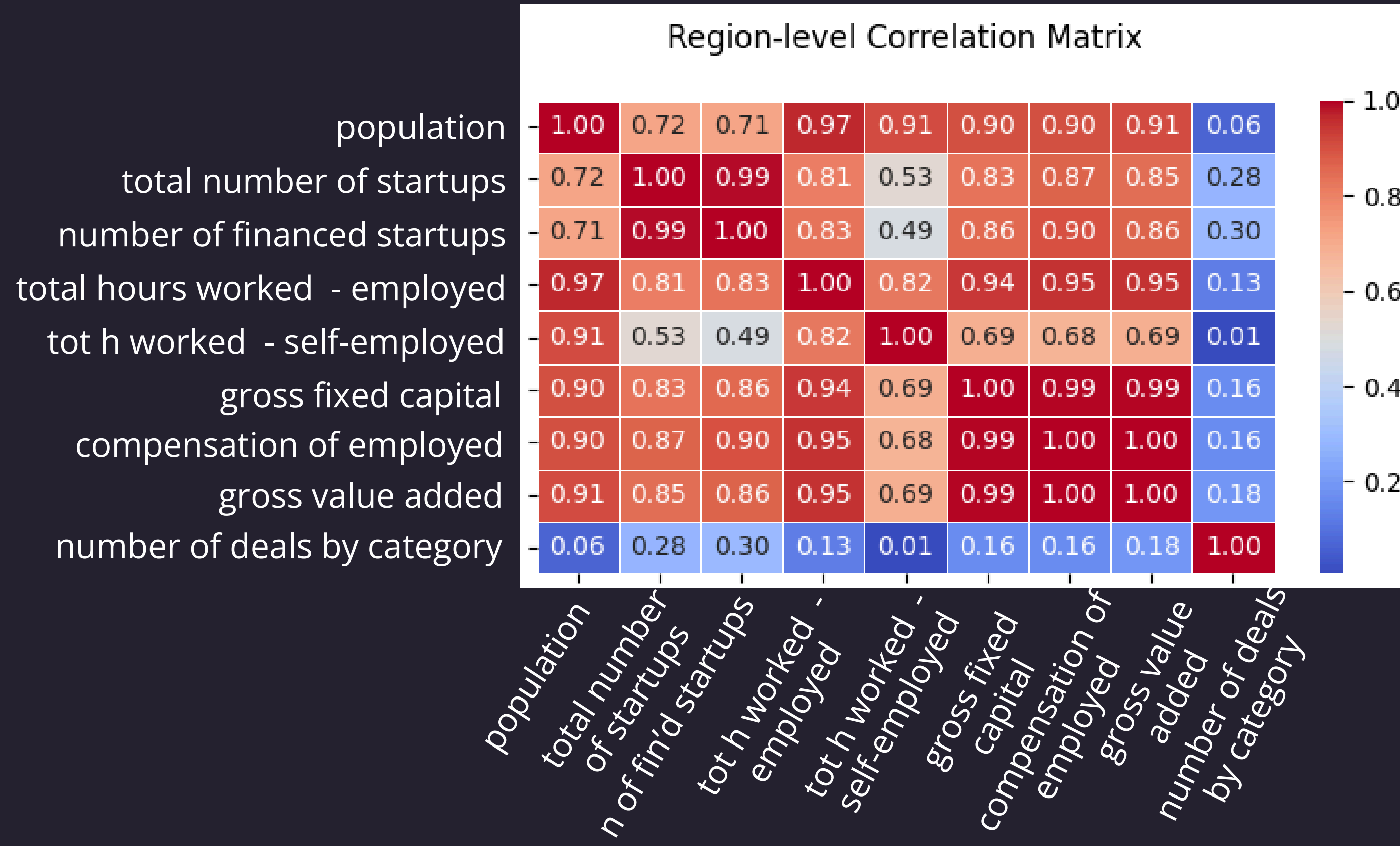
The number of startups financed is growing across Austria after year 2020 and are no longer limited to Vienna alone



# Data analysis



## Correlation with socio-economic variables



# Data analysis



Growth rate of the number of deals by category during the post-pandemic years:

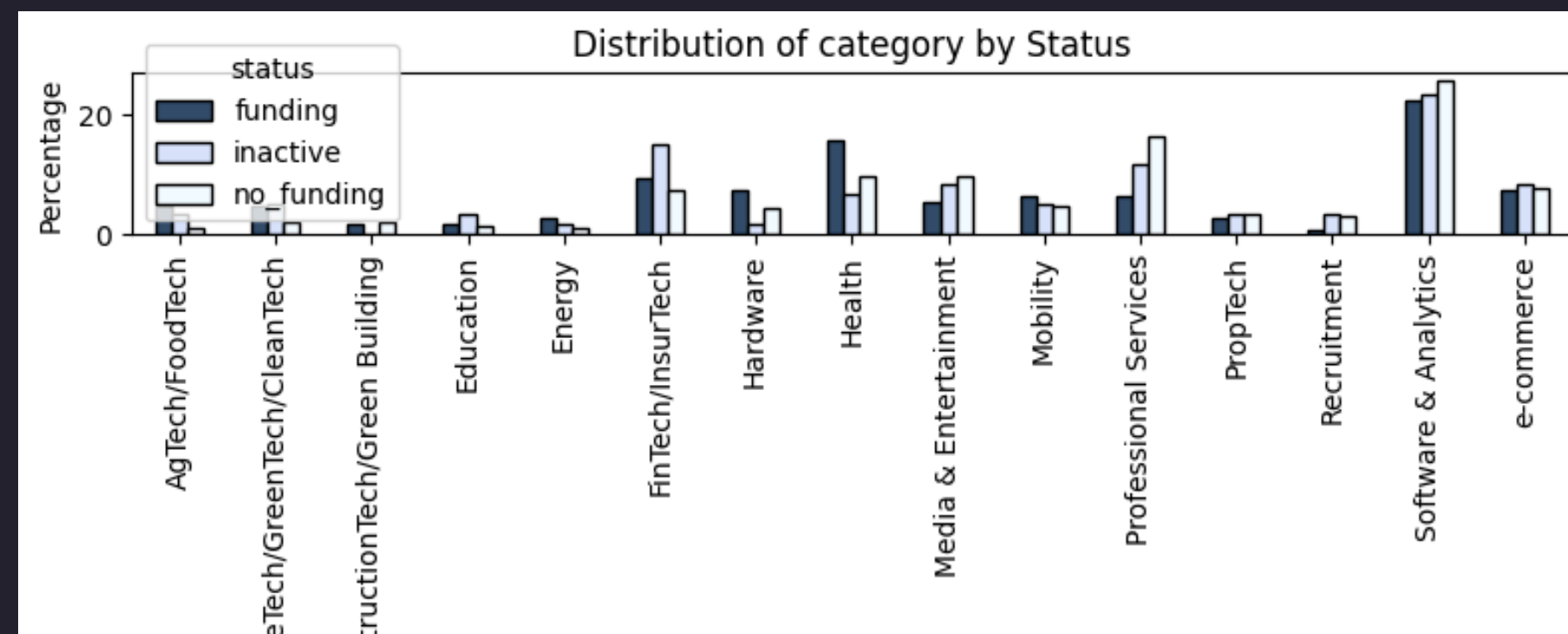
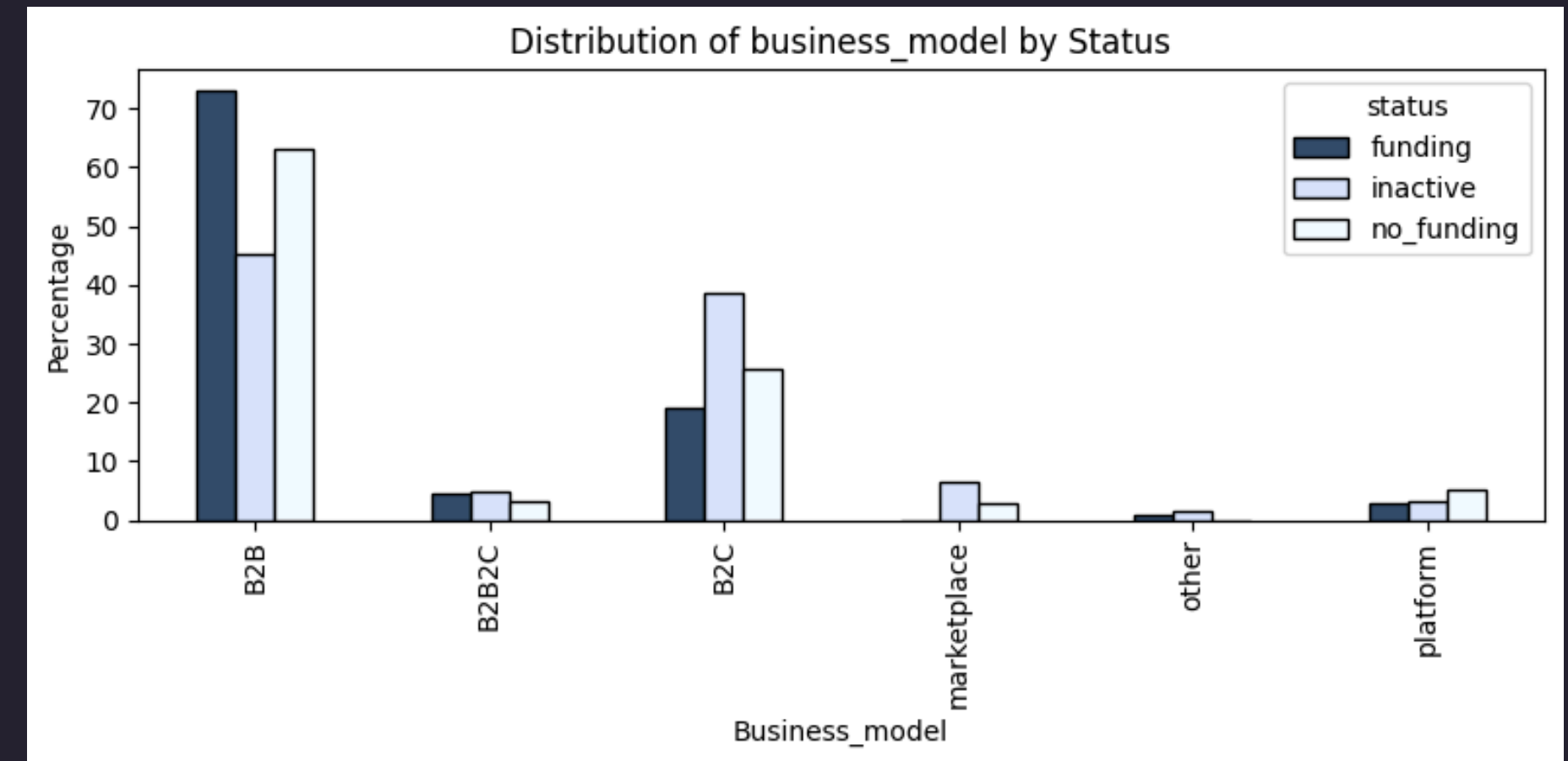
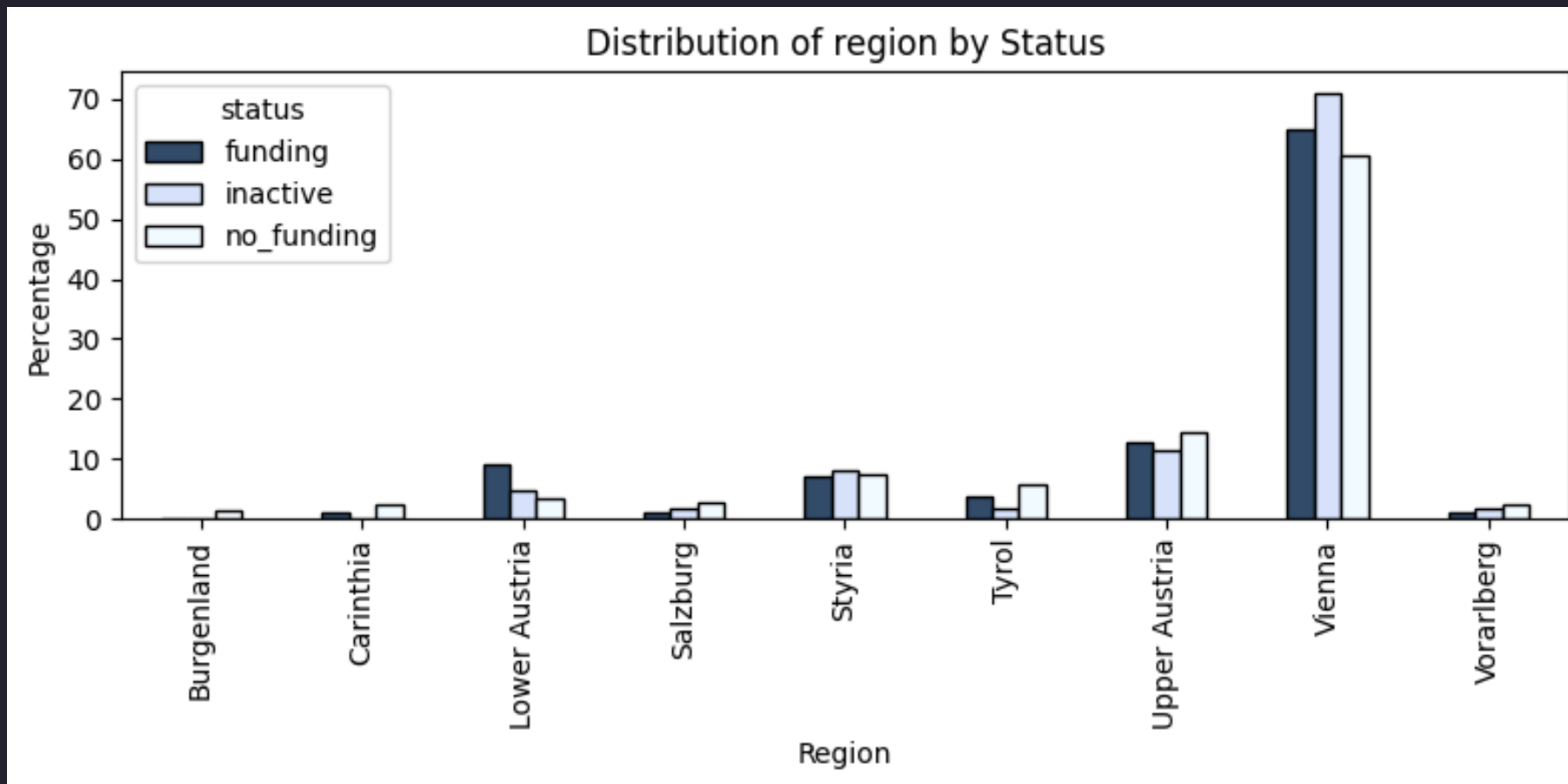
category	growth_rate_deals_by_sector (%)
Software & Analytics	960
Health	375
e-commerce	333
Energy	200
FinTech/InsurTech	120
Mobility	50
ClimateTech/GreenTech/CleanTech	25
Education	20

category	growth_rate_deals_by_sector (%)
ConstructionTech/Green Building	0
Recruitment	0
AgTech/FoodTech	-25
Hardware	-38
Media & Entertainment	-40
Professional Services	-40
PropTech	-71

# Data analysis



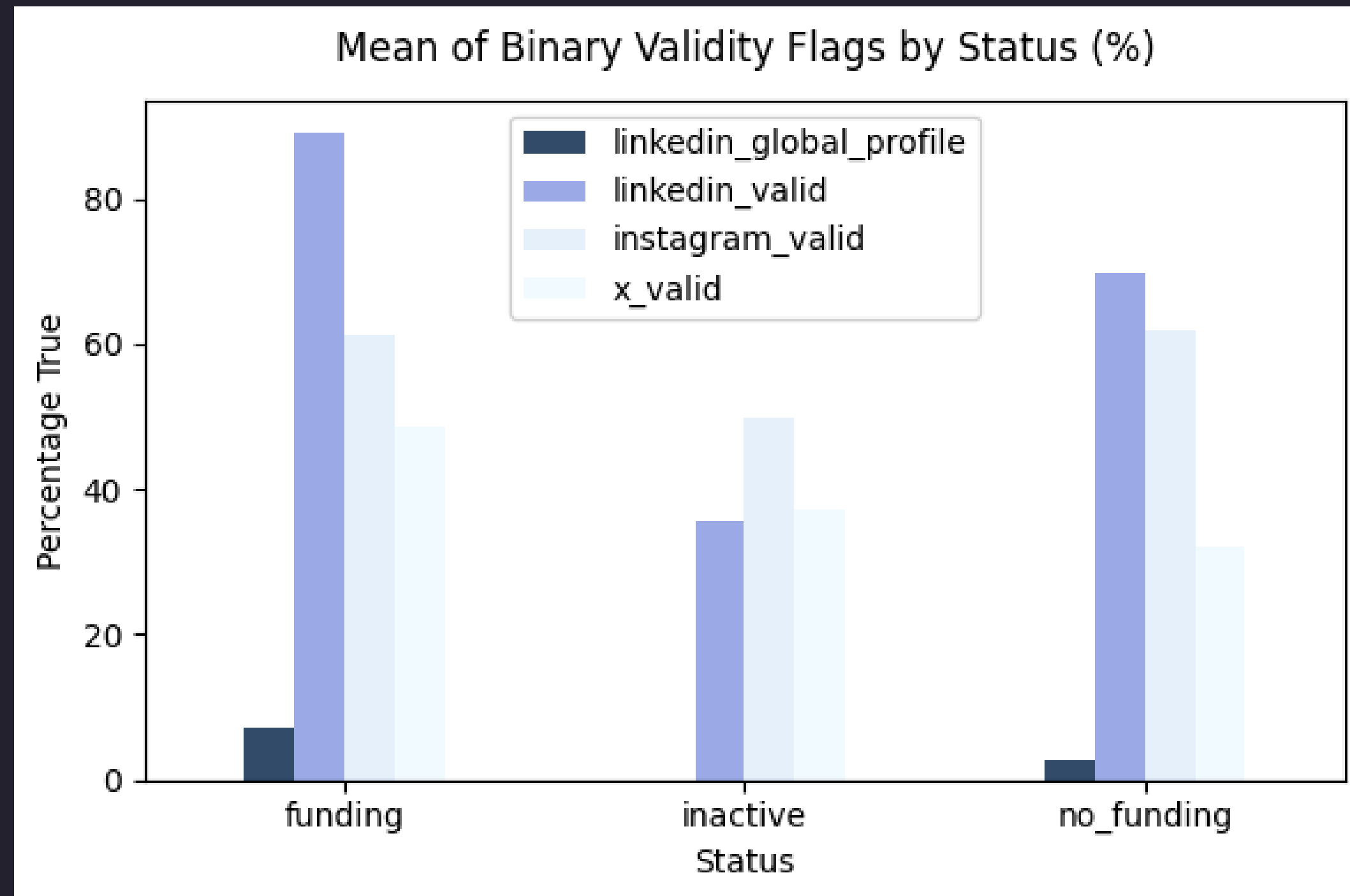
## Categorical variables: percentage distribution by status



# Data analysis



Binary flags: mean true rate by status





# Predictive Model



## Random Forest Status Classification

### Features

- Use Features as predictors:
  - Age
  - Headline Count
  - Description Scores
  - Region
  - Category
  - Business Model
  - Social Media Flag
    - LinkedIn
    - Instagram
    - X

### Model

- Random Forest Model
- Preprocessing Pipeline
- Cross Validation + Grid Search

### Predict Label

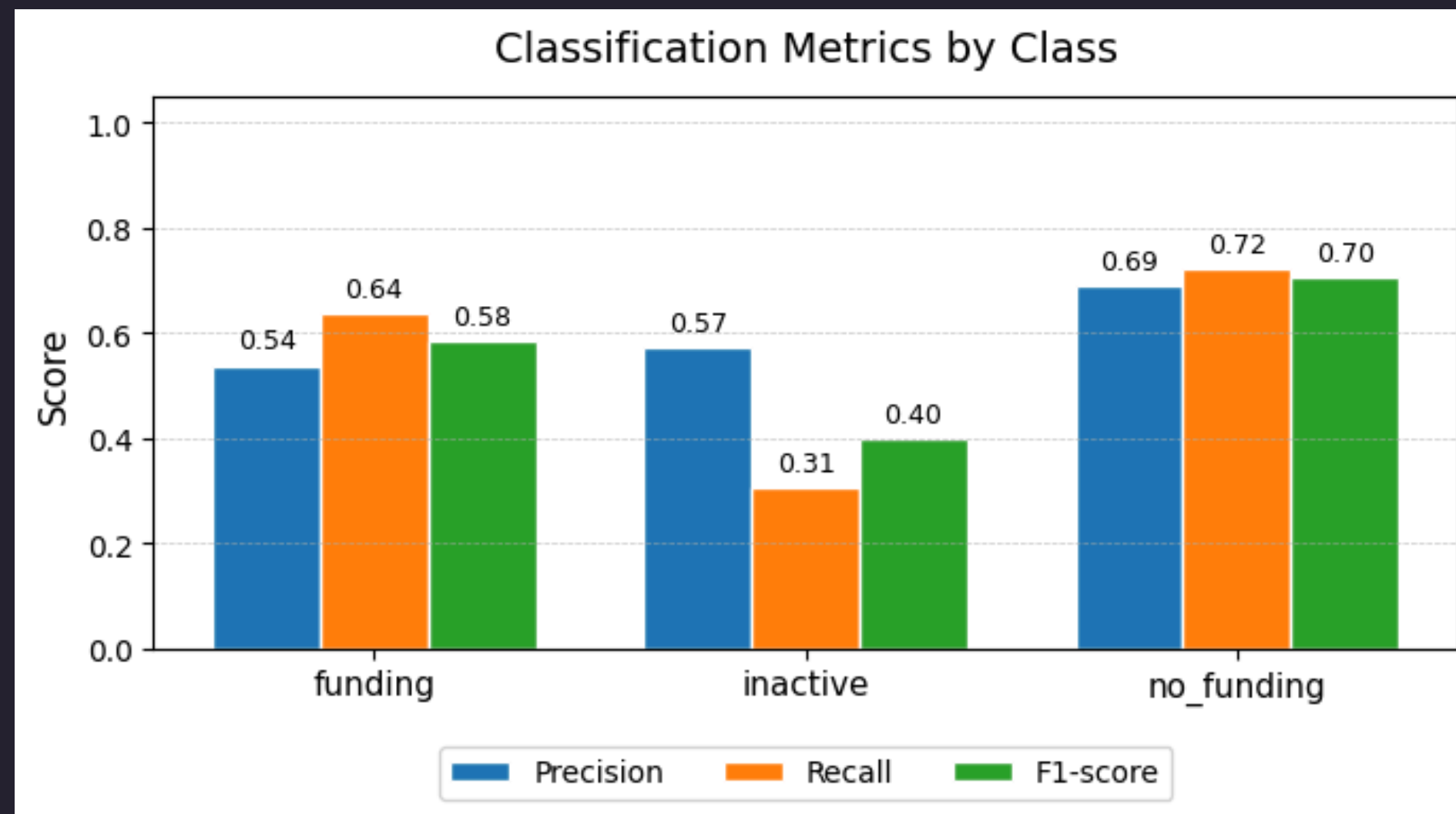
- Prediction on label
  - inactive = startup does not exist anymore
  - no\_funding = startup exists, but got no funding
  - funding = startup exists and got funding

# Model Results 1

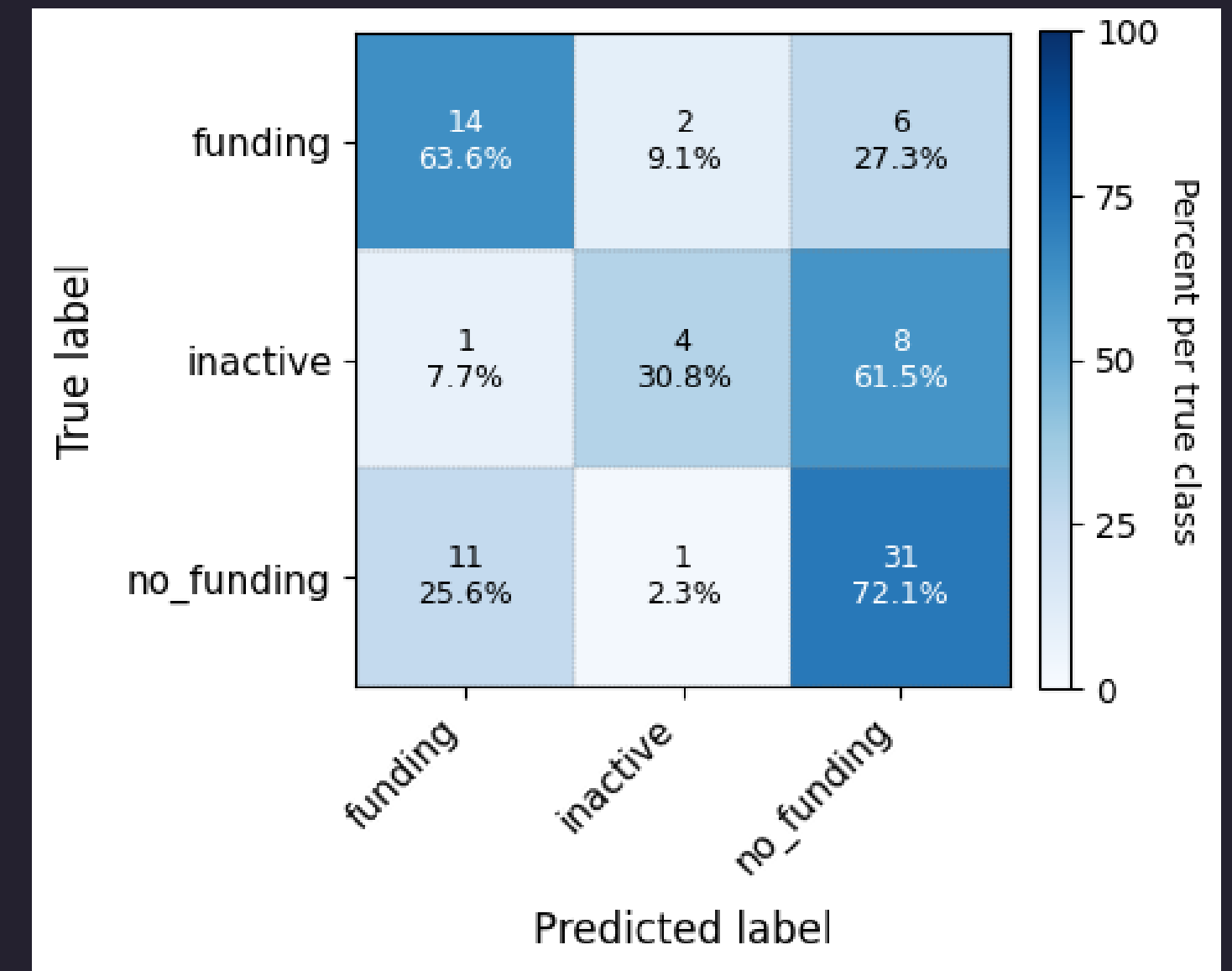


## Accuracy / Metrics:

Overall Accuracy: 63%



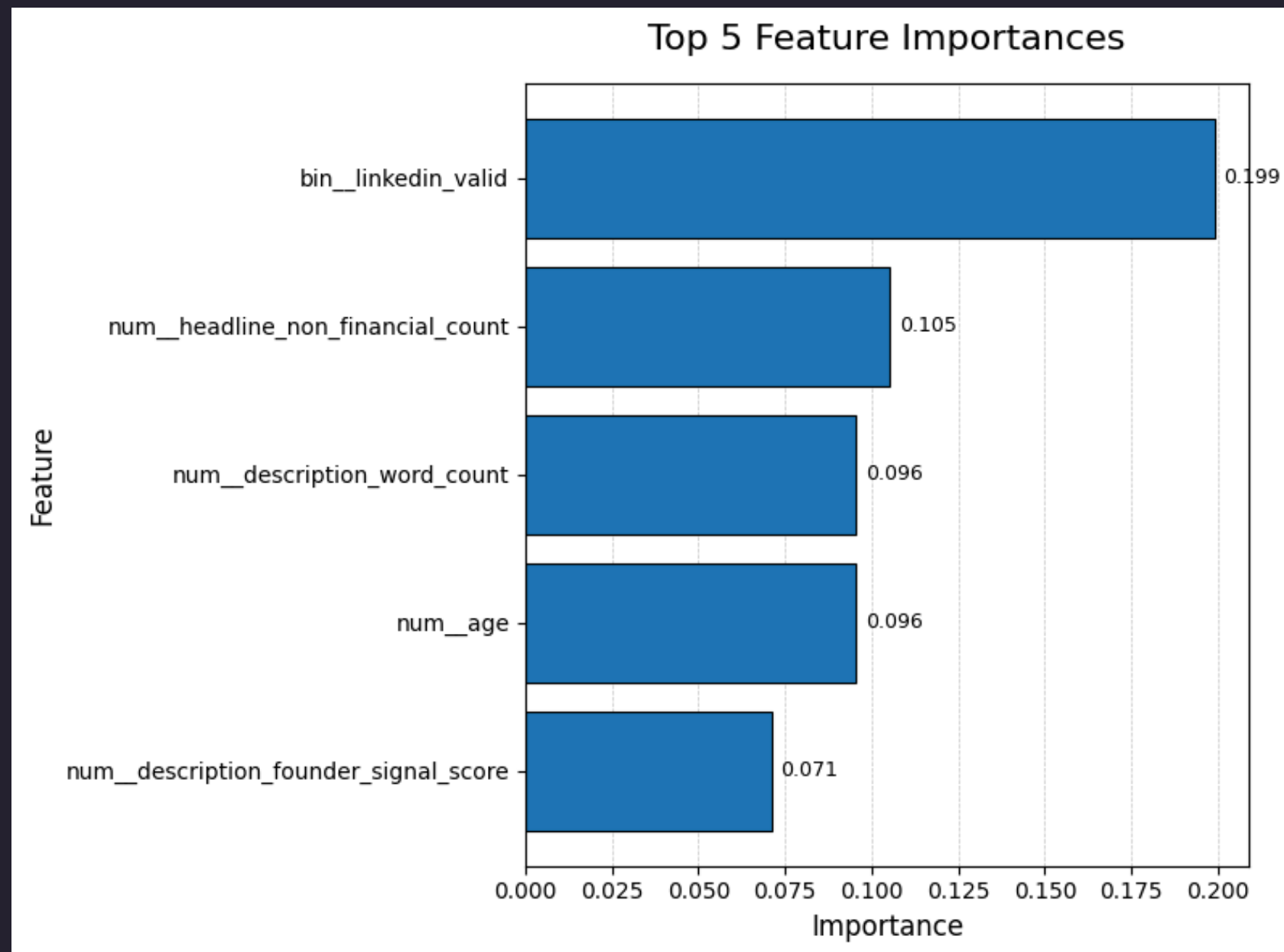
## Confusion Matrix:



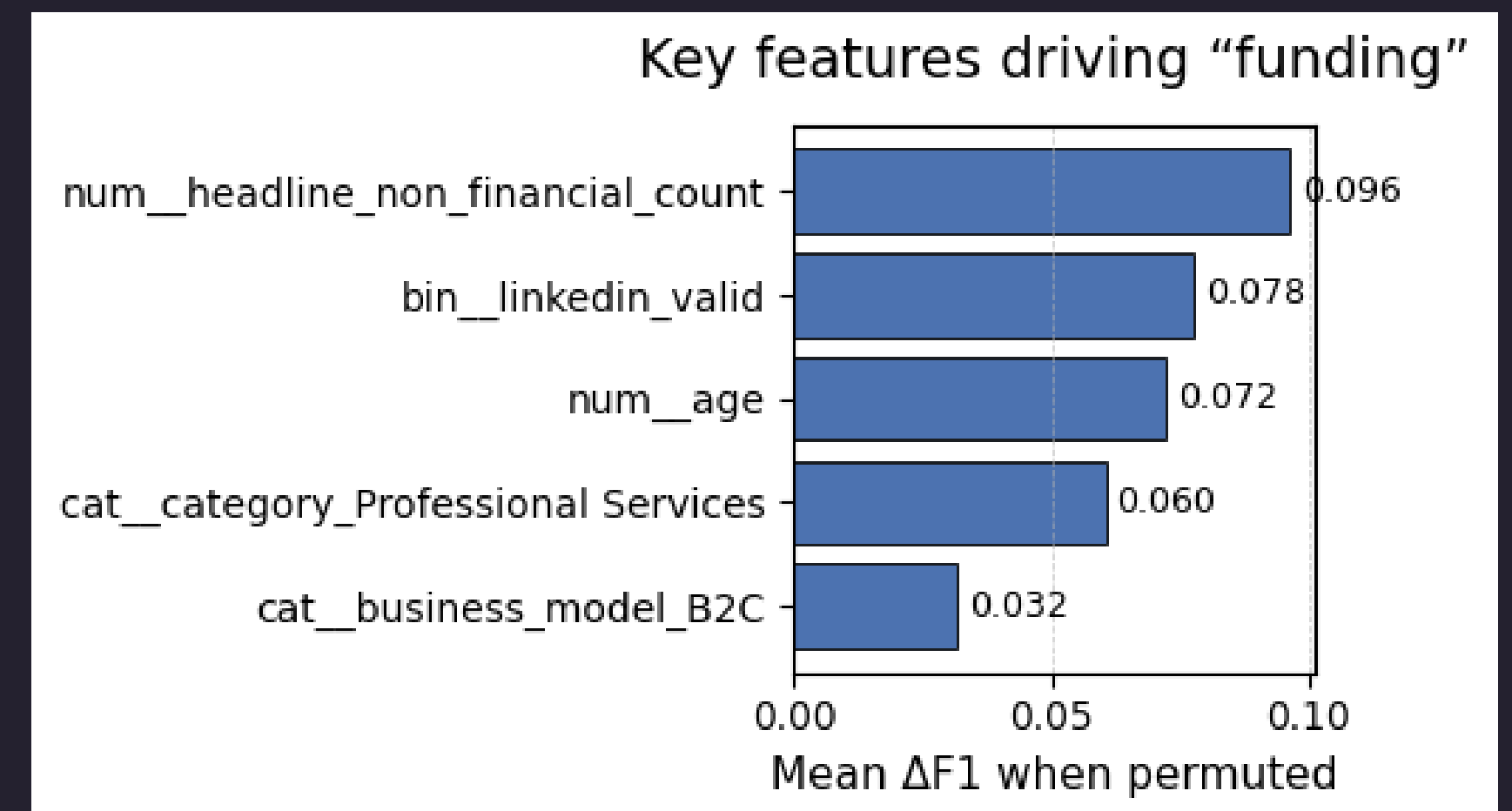
# Model Results 2



## Feature Importance Overall:



## Feature Importance (Funding):



# Final Overview



## What we achieved:

### Deliverables

- Dataset of 400+ startups
- Structured overview of the Austrian startup market
- Predictive model to classify startup funding potential

### Time Split

- 20% — Researching & Ideas
- 50% — Scraping & Data Preparation
- 15% — Data analysis
- 15% — Modeling & evaluation

### Improvements

- Test other models (e.g. XGBoost, LightGBM)
- Add more features
  - (e.g. website traffic, team composition, investor data)