

# IST 782 Applied Portfolio

Francisco Franco Arenas

2024-08-06

# Introduction

- ▶ This is a presentation about the final project of the course IST 782 Applied Portfolio.
- ▶ 3 projects presented:
  - ▶ Project 1: IST 719 Information Visualization: Poster on transitions from University to employment in Spain
  - ▶ Project 2: IST 718 Big Data: ML classifier of financial complaints
  - ▶ Project 3: IST 737 Visual Analytic Dashboards: Europe's quality of life visualization dashboard
- ▶ These 3 projects are the culmination of the learning objectives of the Applied Data Science program at Syracuse University.

## About me

- ▶ Francisco Franco Arenas
- ▶ Industrial engineer with a experience in economics
- ▶ Economist at the Ministry of Economic Affairs in Spain since 2019
- ▶ Started the MS in Applied Data Science at Syracuse University in autumn 2023
- ▶ Passionate about data analysis and visualization

# Why Data Science?

- ▶ Data science stands at the intersection of statistical analysis, computational techniques, and domain-specific knowledge, offering powerful tools for extracting meaningful insights from vast datasets.
- ▶ As an economist at the Ministry of Economic Affairs in Spain, I recognized the transformative potential of data science to enhance decision-making and policy formulation.
- ▶ Pursuing a Master of Science in Applied Data Science was a strategic decision to augment my expertise with advanced data analysis skills.
- ▶ This program has equipped me with the ability to harness complex data, apply sophisticated machine learning algorithms, and develop predictive models, thereby improving my capacity to analyze economic trends, evaluate policies, and make data-driven recommendations.

## Learning Objectives

The main learning goals of the program are as follows:

1. Collect, store, and access data by identifying and leveraging applicable technologies
2. Create actionable insight across a range of contexts (e.g. societal, business, political), using data and the full data science life cycle
3. Apply visualization and predictive models to help generate actionable insight
4. Use programming languages such as R and Python to support the generation of actionable insight
5. Communicate insights gained via visualization and analytics to a broad range of audiences (including project sponsors and technical team leads)
6. Apply ethics in the development, use and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy)

The projects presented in this paper demonstrate how I have met the learning objectives.

# Project 1: IST 719 Information Visualization: Poster on transitions from University to employment in Spain

- ▶ **Objective:** Create a poster that visualizes the transitions from University to employment in Spain.
- ▶ **Dataset:** University Placement Survey of Spain
- ▶ **Tools:** R, ggplot2, Adobe Illustrator
- ▶ **Techniques:** Alluvial plot, Ridgeline plot, Cleveland dot plot, etc.
- ▶ **Learning:** Advanced visualization techniques, effective data communication

## Project 2: IST 718 Big Data: ML classifier of financial complaints

- ▶ **Objective:** Develop a machine learning classifier to predict the outcome of financial complaints.
- ▶ **Dataset:** Consumer Financial Protection Bureau (CFPB) complaints
- ▶ **Tools:** PySpark, Google Colab
- ▶ **Techniques:** Text preprocessing, feature extraction, machine learning
- ▶ **Learning:** Big data processing, machine learning on large datasets, text classification

## Project 3: IST 737 Visual Analytic Dashboards: Europe's quality of life visualization dashboard

- ▶ **Objective:** Develop an interactive dashboard to visualize the quality of life in the European Union.
- ▶ **Dataset:** Eurostat data on quality of life indicators
- ▶ **Tools:** Tableau, R
- ▶ **Techniques:** Data collection, data preparation, interactive dashboard design
- ▶ **Learning:** Data visualization, interactive dashboards, storytelling with data