# Twitter ppt

## Introduction

* Description of the **problem**
* Research **motive**/question
* **Resource and limitations**
  + Twitter API
  + VM

## Data preparation

* **Accounts** (generalized, 15 categories, 15 aprox per account)
* **Data statistics** (timeframe, collected tweets, volume of data)
* **Data exploration** (data cleaning, graphs)

## Data Analysis

* **Data set structure** 
  + TABLE: (target variable and predictors/features)
  + GRAPH: Clusters
* **Linear regression** (add regression hypothesis graphs)
  + Only followers
  + All
* **Polynomial regression**
* **Regression tree** (graph)

## Conclusions and next steps

* **Graphs**
  + **Geographical Map** (missing account locations, localized data -> biased)
  + **World Cloud** (1 day, Sunday the 22nd for all accounts and divided by category, FASHION)
* **Conclusions and lessons learned**
  + (no substantial relation daily activity/followers, our data sample is not representative of Twitter)
  + Possible different data collection and type (dictionary), be more careful with text encoding (UTF-8), social media analysis is hard
* **Future lines of work**
  + Time series analysis
  + Social Network analysis (node centrality, reach, etc)
  + Followers heat map
  + Wider range and variety of data
  + NLP

## Model deployment

* Graphs with ALL
* Live code
* Weird observations