EA Test

October 2019

Environment setting

❑ Anaconda distribution for Win10

❑ Installing necessary libraries for database deployment & connection + interface

➢ PostgreSql

- conda install -y -c conda-forge ipython-sql
- conda install -c anaconda postgresql
- conda install -y -c anaconda psycopg2

➢ MongoDB

- conda install -c anaconda mongodb
- conda install -c anaconda pymongo

❑ Installing necessary library for accessing Twitter API
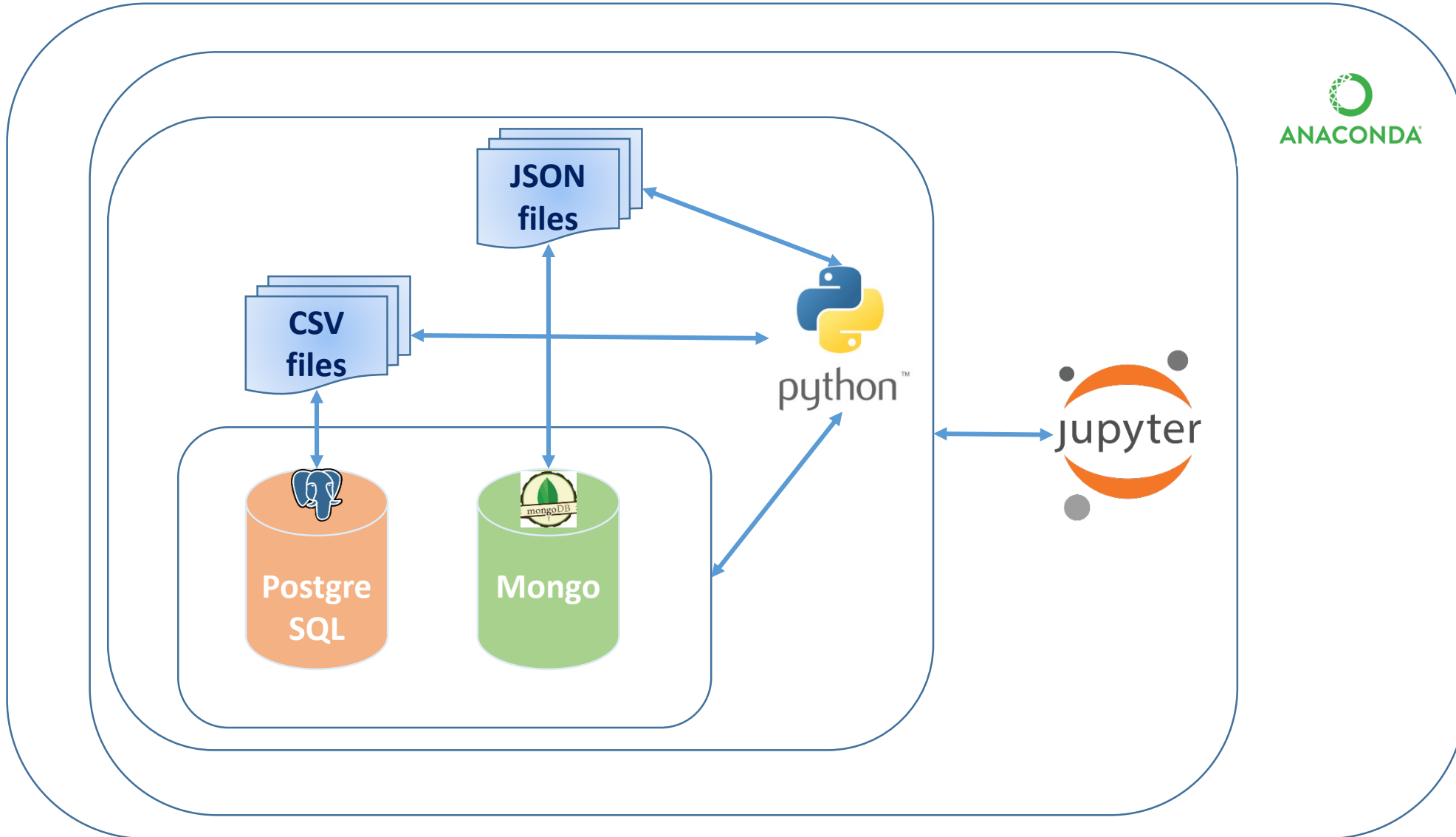- conda install -c conda-forge tweepy

Database Setup & Start steps

❑ PostgreSql from Conda prompt :

  ➢ set PGDATA=C:\Users\iteofil1\Desktop\EA\postgreDB
  ➢ mkdir %PGDATA%
  ➢ pg_ctl initdb
  ➢ pg_ctl start
  ➢ psql -c "create database EA_TEST;" postgres

❑ MongoDB from Conda prompt :

  ➢ mongod --dbpath "C:\Users\iteofil1\Desktop\EA\mongoDB"

Architecture diagram

Tackling the objectives

❑ Task 1 Uploading info into database

Handling structured (csv) and unstructured (json) data implies thinking in different databases.

For **structured** data **relational Database** performs nice, and from open source solutions I chose PostgreSQL.

Unstructured data need **NoSql Database**, and MongoDB is fast and widely used in open source solutions.

❑ Task 2 Data quality : Analyze the quality of the extracted data and transform date if it´s necessary.

Pandas is a powerful tool when dealing with data quality. It allows you to explore data formats, to transform from a type to another, to filter nulls, etc

Pandas also give us the capabilities of merging and joining datasets.

❑ Task 3 EDA Exploratory data analysis of the FIFA 2018 Player dataset that the main analysis will be the different languages of the players speak.

  Also for this point Pandas is a great option. You may summarize, aggregate, etc.

  Importing seaborn and matplotlib is necessary for plotting
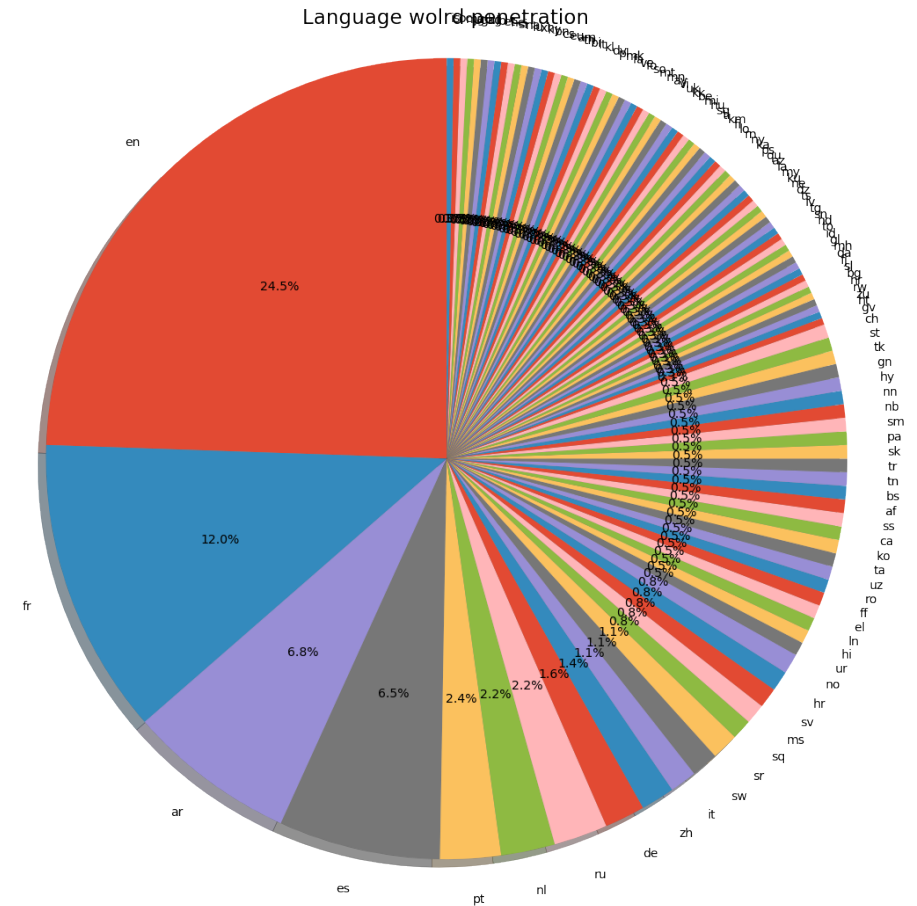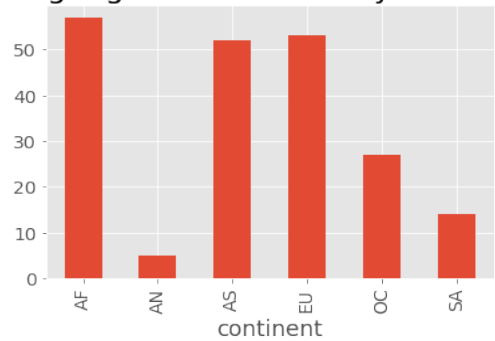
❑ Task 4 : Data visualization

  Cluster plots help understand the information for data driven decissioning

❑ Task 5 : Sentiment analysis. Extract information on social media to obtain information about the new game FIFA 20
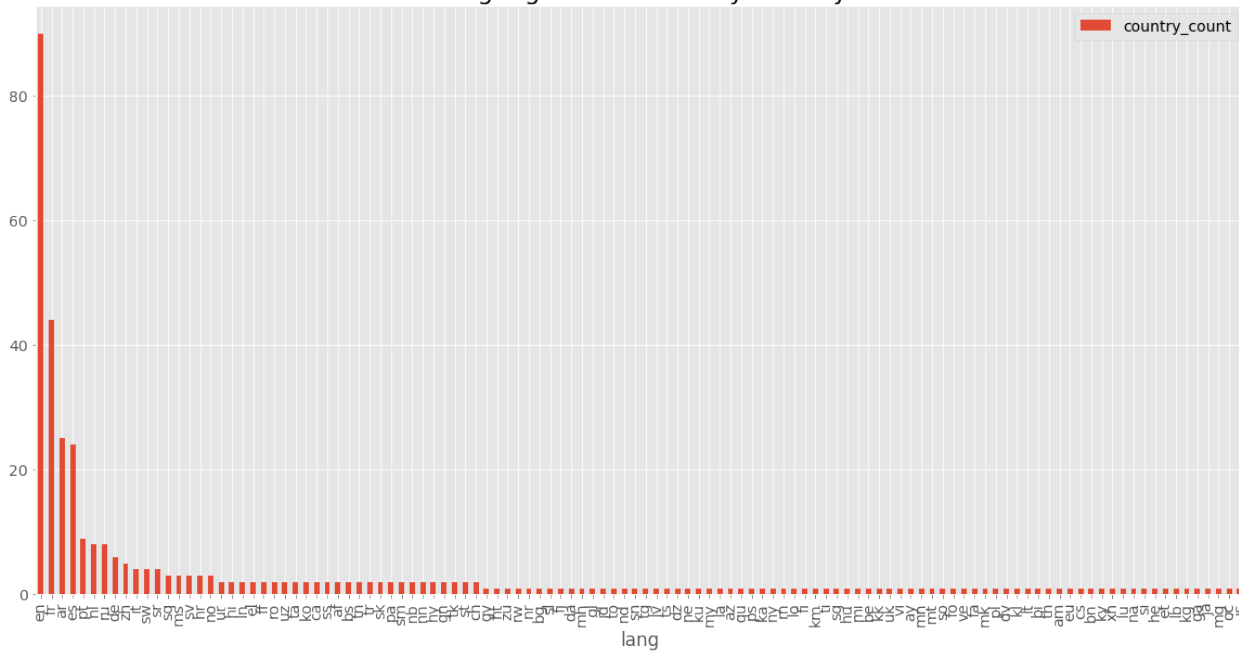
Business Insigths : localization

There are 12 languages spoken in more than 5 countries around the wolrd

Languages distribution by continent

Language wolrd penetration

Business Insigths : localization

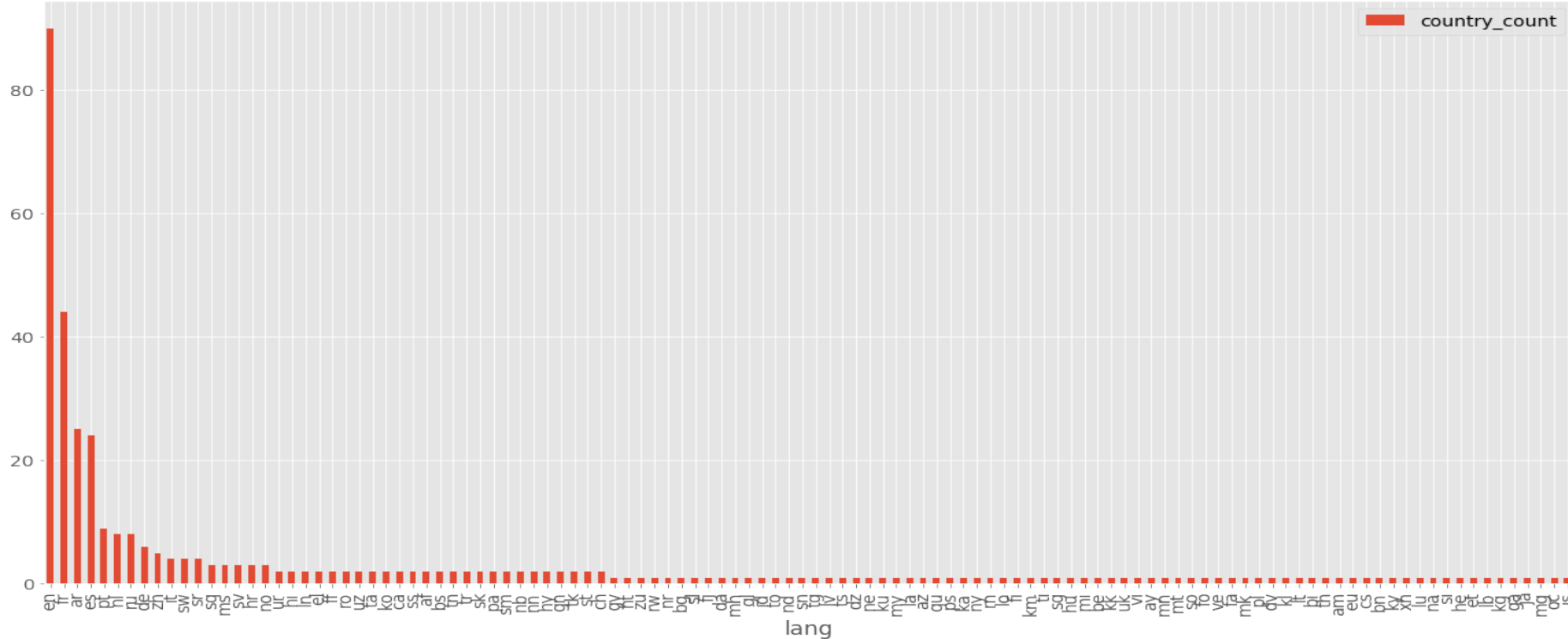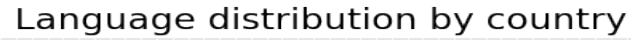Language distribution by country

There are 12 languages spoken in more than 5 countries around the wolrd

- ✓ English is   spoken in more than 90 countries
- ✓ French is   spoken in more than 40 countries
- ✓ Arabic is   spoken in more than 20 countries
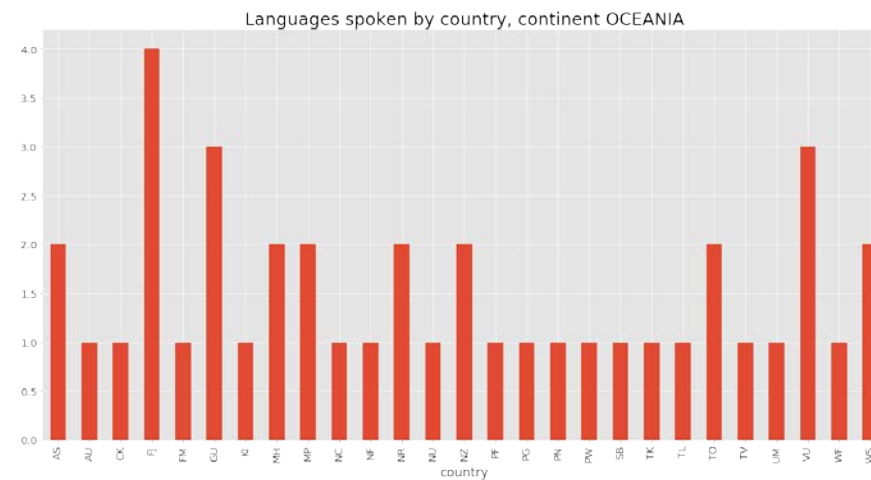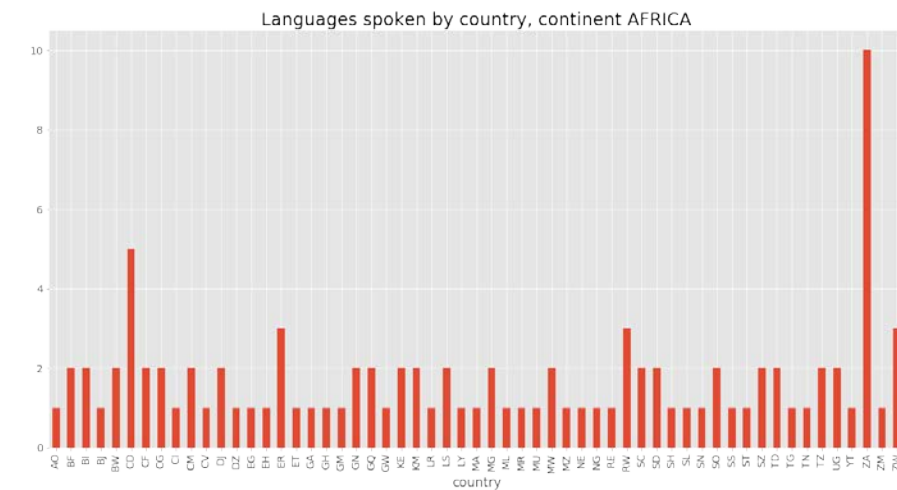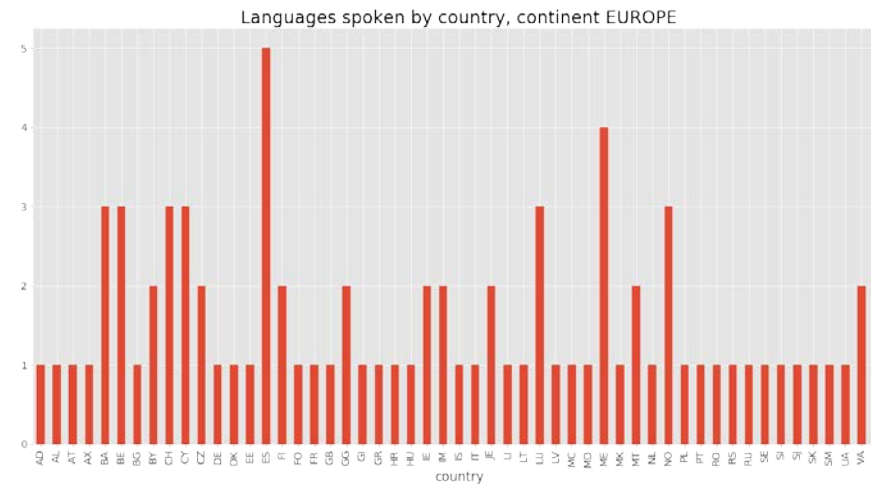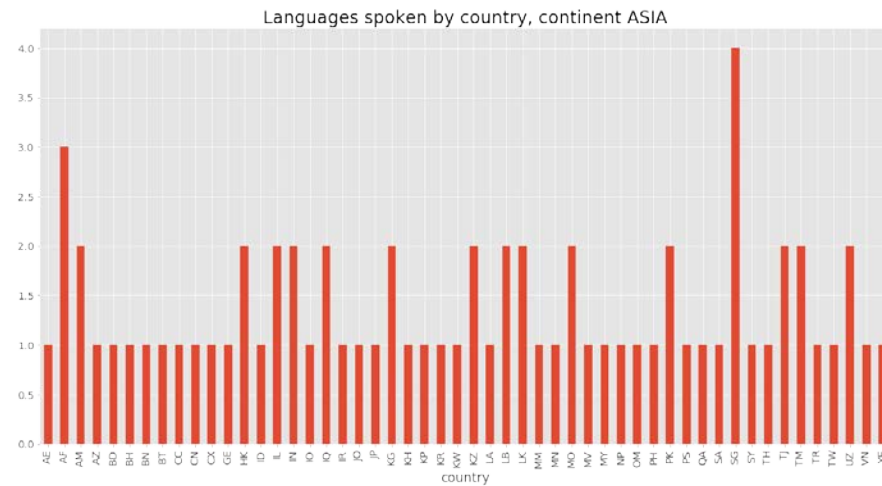- ✓ Spanish is   spoken in more than 20 countries

Business Insigths : localization

Language distribution by country

- ✓ English is spoken in more than 90 countries
- ✓ French is spoken in more than 40 countries
- ✓ Arabic is spoken in more than 20 countries
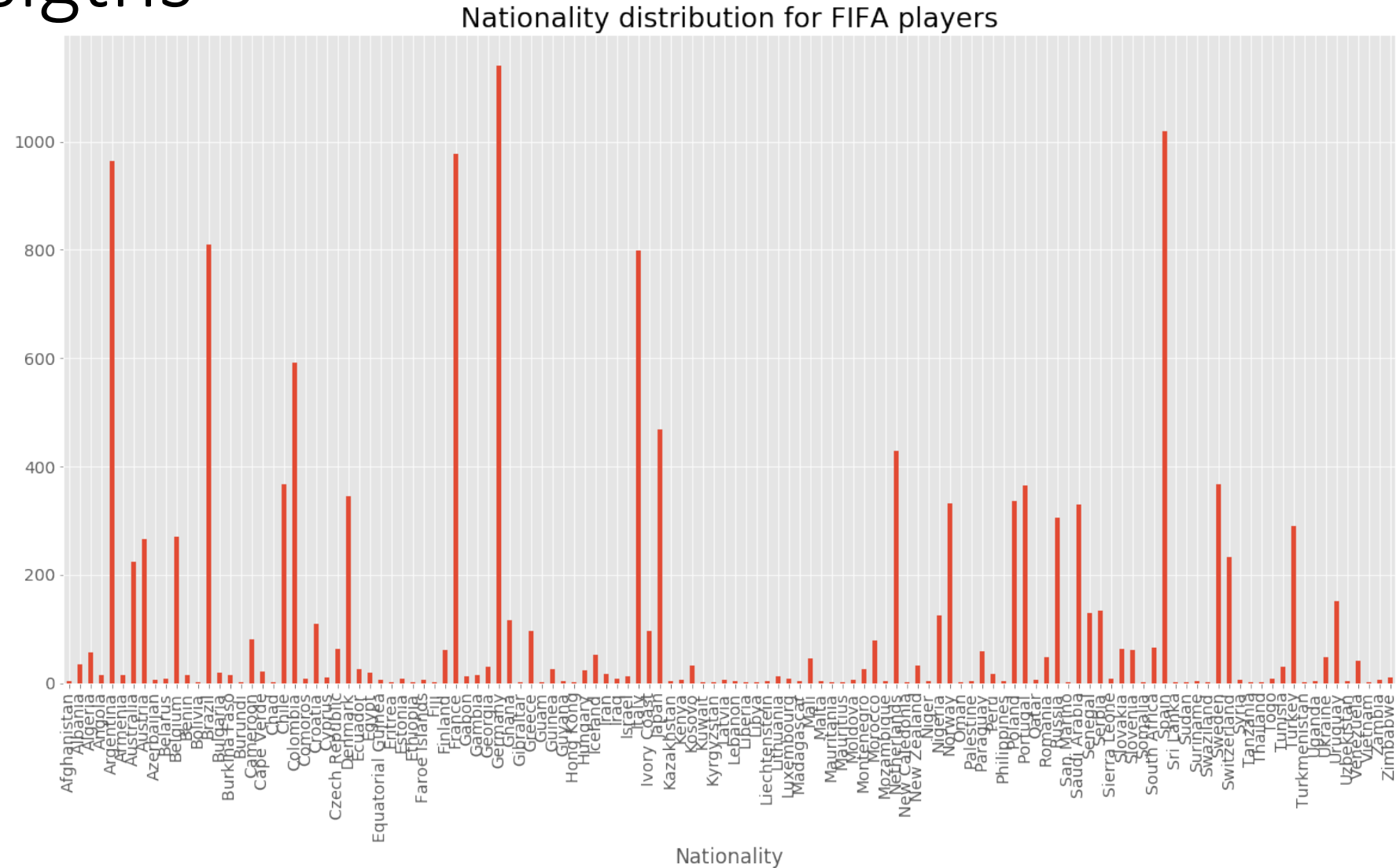- ✓ Spanish is spoken in more than 20 countries

Business Insigths : localization

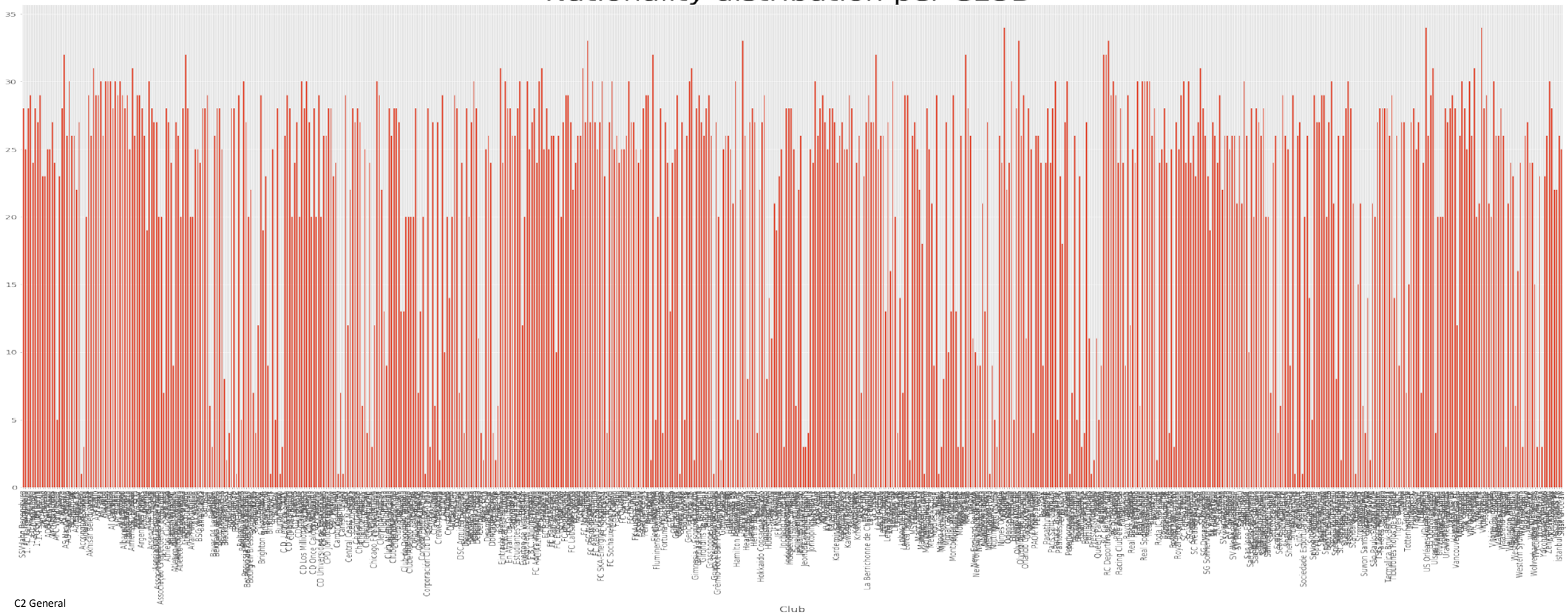✓ Languages spoken by country per continent

Business Insigths

✓ Less than 10 nationalities represent half of total players

✓ Classical football countries are well identified by this graph

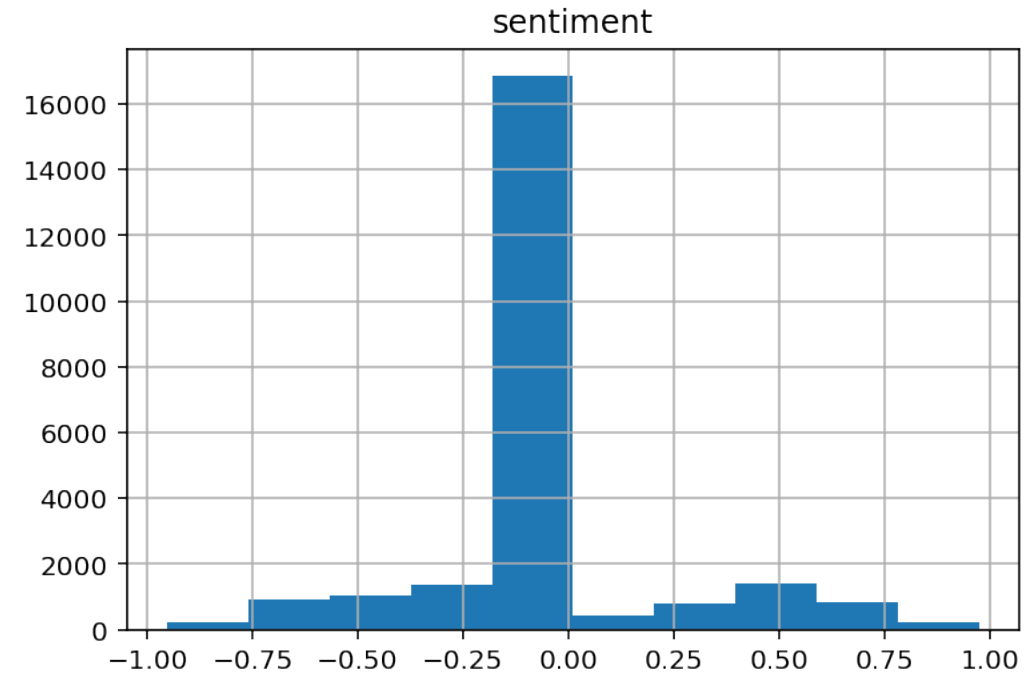Nationality distribution for FIFA players

Business Insigths

✓ Looking at Club level, there are so many nationalities represented that universal language is a must depending on continent

Nationality distribution per CLUB

Business Insigths : sentiment analysis

✓ Based on Tweets, searched by
  FIFA20. 25K sample size

sentiment

Business Insigths : sentiment analysis

✓ Postive sentiment relevant
  topics

✓ Negative sentiment relevant
  topics

Thank you !