Gimnazija Jovan Jovanović Zmaj

Novi Sad

Maturski Rad

Predmet:

Objektno orijentisano programiranje

Tema:

Obrada i grafički prikaz podataka o kafi svetskog uzgoja

Mentor: Učenik:

Stanković Milan Ekres Alek

Novi Sad, Januar 2021

Sadržaj

No table of contents entries found.

# Uvod i obrazloženje

Cilj ovog maturskog rada je dvojake prirode. Sa programerskog aspekta cilj je uskladiti različite alate da bi se postigao cilj obrade i prikaza podataka koristeći programski jezik python, a sa druge strane, čitajući podatke, dođemo do zaključka: koje su bile pogodne godine za uzgoj kafe i koje države proizvode najkvalitetniju kafu, takođe i upoznavanje sa parametrima uz čiju pomoć dolazimo do tog zaključka.

Kroz ovaj rad

# Izrada zadatka

## Obrada Podataka

Počinjem tako što preuzmemo podatke sa sajta

pošto su preuzeti podaci u tekstualnom formatu potrebno je da ih prebacimo u format koji nam može koristiti pri obradi (.xlsx). Preko data panel-a koristimo opciju from text formatiramo .csv file kako bi ga koristili. Od svih podataka koje posedujemo pola necemo koristiti i zato ih brišemo u ovom koraku. Neke od kolona koje ćemo koristiti su: “Place”, “Country of Origin”, “Region”, “Distributor”, …

Sledeći korak je očistiti podatke koje ćemo dalje koristiti. Proces “čišćenja” se svodi na formatiranje i manipulisanje podacima tako da nemamo praznih podataka, podatka koji nisu u istom formatu kao I drugi podaci iste kolone I podataka koji su slučajno premešteni. Za to koristimo python biblioteku pandas. Pandas je open source bilblioteka od 2009 i koristi se nasiroko u radu sa podacima jer cini lakim sve randnje koje se tiču dataframe-a. Dataframe je pandas objekat koji pored svoje brzine i efikasnosti obrade poseduje integrisano indeksiranje za lakoću rada, pregled podataka i kompatibilnost sa standardom relacionih sql baza podataka. Kasnije iz toga sledi lak rad sa prethodno spomenutom bazom što bi inače predstvljalo problem zbog kompatibilnosti podataka, preuzimanjem istih i kasniom upotrebom. Kako se može koristiti za skladištenje podataka i prikaz, sa istom lakoćom, sa svega nekoliko linija koda, možemo iterisati sve naše podatke i čistiti ih. Za taj zadatak koristili smo jupiter notebook koji zajedno sa anacond-om može obezbediti virtuelno okruženje i detaljan pregled koda, kao i mnoge druge alate i korisne funkcije. Značajniji problemi su format datuma ocenjivanja (pr. March 2nd, 2010) i godine uzgoja koje su date kao period (pr. 2014-2015). Za godine nam je potrebna jedna godina tipa integer a što se tiče datuma moramo ga prebaciti u format date sql baze %Y-%M-%D (pr. 2010-03-02). Tako očišćene podatke skladištimo nazad u .csv file kako bi ih importovali u bazu(slika).

## Rad sa bazom podataka

Mysql je najpopularnija open source baza podataka. Sa nizom godina dokazanim performansama i lakoćom korišćenja postala je glavni izbor za web-based aplikacije.Mysql pripada firmi Oracle i danas mysql koriste i najveće firme visokog profila kao što su Facebook, Twitter, Youtube i Yahoo. Mysql se danas takodje koristi za pravljenje cloud-based aplikacija koje su sposobne da čuvaju velike količine informacija uporedjenju sa prethodnim generacijama, što dodatno čini mysql tehnologiju pouzdanim izborom većine programera. Zbog lakoće korišćenja i efikasnosti mysql je naš izbor. Za početak potrebno je osposobiti server kasnije na napravljenom serveru napraviti shemu i tabelu u shemi i kasnije uneti podatke iz .csv file-a. Ove zadatke mnogo olakšava mysql workbench. Pored toga što sadrži detaljno odradjen grafički interface, za rad nad bazama nije potrebno otkucati ijednu liniju sql koda koji može izgledati zastrašujuće pogotovo ljudima koji se susreću sa sql-om prvi put. Tabelu i podatke uneti preko import wizard-a koji se može pronaći desnim klikom na opciju: “Tables” unutar sheme.

## Anaconda okruženje

Anaconda je platforma za data science; tačnije za ljude koji se bave: podacima, njihovim efikasnim i sigurnim skladištenjem, prikazom, čitanjem i obradom. Anacondu smo takodje koristili u prethodnom koraku da pokrenemo pandas ali nam nije bila toliko neophodna koliko u ovom koraku. Za upravljanje dash aplikaciom biće nam potrebno virtuelno okruženje koje sa čijim radom nam olakšava anaconda. U anaconda client-u sa lakoćom je moguće instalirati sve potrebne datoteke i pakete koji će nam biti potrebni za dalji rad. Potrebno je napraviti okruženje i u njemu instalirati dash (i njegove “core” komponente), datoteke potrebne za rad u jupiter lab-u, pandas, plotly i druge navedene u prilogu. Programiranje obavljamo u jupiter lab-u i njemu možemo pristupiti direktno iz anaconda client-a.

## Dash layout

Dash je python framework za pravljenje analitičkih web aplikacija. Pošto je baziran na Flask, Plotly i React arhitekturi omogućava lak i intuitivan rad sa podacima u python-u. Kroz objekat app pokrećemo preko par ugradjenih funkcija (prilog). Ostale elemente koje unosimo jesu html elementi u koje unosimo grafike i tekst koji želimo prikazati na stranici. Dash u pozadini te elemente prebacuje u html i css kod a grafike formira preko callback-ova i plotly bibliotekegfg

# Literatura

1. <https://www.kaggle.com/ankurchavda/coffee-beans-reviews-by-coffee-quality-institute>
2. <https://pandas.pydata.org/about/index.html>
3. <https://www.mysql.com/about/>
4. <https://www.anaconda.com/about-us>