# Документација за проектот

### **Краток опис:**

Овој проект опфаќа собирање на податоци од веб сајтот на [CoinMarketCap](https://coinmarketcap.com/). Составен е од две скрипти, и тоа:

* Со едната скрипта се собираат податоци за промената на вредноста на некоја крипто валута на дневно ниво. Податоците собрани со оваа скрипта можат да се најдат во фолдерот [Scraped History Data](https://drive.google.com/drive/folders/1s_0Jdt1lA14pLpGAcNfda1cs-ElEEdG5).
* Со другата скрипта се собираат вестите за крипто валутите од агрегаторот на вести на CoinMarketCap. Податоците собрани со оваа скрипта можат да се најдат во фолдерот [Scraped News](https://drive.google.com/drive/folders/100Q0fL4JYAQQAjHH0CN9cxOlscIrmGW6).

### **Користени технологии:**

За изработка на проектот е користен програмскиот јазик [Python](https://www.python.org/). За собирање на податоците се користи библиотеката [Selenium](https://www.selenium.dev/), додека за екстракција на податоците во .csv фајл се користи библиотеката [pandas](https://pandas.pydata.org/).

### **Користени ресури за учење:**

На почетокот, изгледав два онлајн курсови за web scraping:

* [Scraping the Web for Fun and Profit](https://jakobgreenfeld.gumroad.com/l/scraping) од Jakob Greenfeld
* [Web Scraping 101](https://www.founderhacker.com/web-scraping) од Ryan Kulp

Овие два курсеви комбинирано навистина беа од помош, најмногу поради тоа што беа целосно практични, односно материјалот беше објаснет преку создавање на проект со use case во вистинскиот свет.

Освен тоа, користев и Youtube туторијали и Stack Overflow прашања кога ќе наидев на некој специфичен проблем.

### **Соочени проблеми**:

Проектот беше околу 80% готов во јуни, но поради други обврски морав да паузирам во јули и август.

Во тој период, имало парцијален редизајн на агрегаторот на вестите на CoinMarketCap, па поради тоа кога продолжив со работа на крајот на август, голем дел од скриптата што беше напишана не работеше поради редизајнот.

Исто така, друг чест проблем беше и лошата интернет конекција, бидејќи скриптата ја пуштав локално.

### **Стартување на проектот:**

Двете скрипти кои се користат за проектот се наоѓаат во фолдерот “SLearning Selenium”.

Пред да се стартува проектот, треба да се специфира локацијата на ChromeDriver-от во деветта линија од двете скрипти.

За стартување на проектот само треба да се одбере за кои крипто вредности сакаме да собираме податоци и потоа само кликаме на “Run” во IDE-то.

За **Scraped History Data** одбирањето на крипто валути се прави со тоа што во низата crypto\_names ќе ги додадеме целосните имиња на крипто валутите за коишто сакаме да собираме податоци.

За **Scraped News** во низата crypto\_names додаваме дел од url-то за дадената крипто валута. Пример: URL адресата на крипто валутата со име Battle World e [https://coinmarketcap.com/currencies/**battle-world**/](https://coinmarketcap.com/currencies/battle-world/), па во овој случај во crypto\_names би додале: “**battle-world**”

### **Output files:**

1. За фајловите од **Scraped History Data**, се собираат следните податоци:
   1. **Day** - денот, месецот и годината за којшто се однесуваат останатите податоци.
   2. **Open** - цената на крипто валутата на почетокот на тој ден.
   3. **High** - највисоката цена што крипто валутата ја достигнала на дадениот ден.
   4. **Low** - најниската цена што крипто валутата ја достигнала на дадениот ден.
   5. **Close** - цената на крипто валутата на крајот на тој ден.
   6. **Volume** - покажува колкава количина од дадената крипто валута се има трејдано на тој ден (во долари)
   7. **Market Cap** - покажува колкава е целосната вредност (во долари) на сите coins што се ископани за таа крипто валута до тој ден.
2. За фајловите од **Scraped News**, се собираат следните податоци:
3. **Headline** - насловот на веста.
4. **Body** - краток preview од целосната вест.
5. **Posted By** - кој веб сајт ја има постирано таа вест.
6. **Time Posted** - во кое време од денот е постирана веста.
7. **Trust Score** - дали веста е вистината или лажна според интерналниот систем гласање на CoinMarketCap каде што секој што има CoinMarketCap може да гласа. Доколку резултатот е Fact/Fud значи дека немало доволно гласови да се добие Trust Score за таа вест.
8. **Crypto Name** - за која крипто валута се однесува веста.