SCRYPTOOL

Cryptocurrency intelligence solution powered by a multi-source scraping tool.

Web Analytics Final Project - Team 2

Sergio Aizcorbe, Bernardo Bouzas, Jaime Mardones, Manuel Nuño

1. Concept Design Process
2. Google Data Studio Dashboard
3. Scraping strategy and implementation
4. Natural Language Processing
5. Recap and potential improvements opportunities
6. Other state-of-the-art solutions
7. Attributions and references

Concept Design Process

1. Concept Design Process

Idea

Scrape, process, combine and analyze **18 cryptocurrencies** data.
Obtain it from different reliable sources and in a **3-month** date range.
Display the result in an **interactive dashboard** that offers insightful visualizations.

Metrics

- Financial statistics
- Social volume
- On-chain data
- Developers activity
- Sentiment analysis
- Word vectorization

Workflow

Design → Research → Scrape → Process → Analyze → Export → Visualize

(If needed)

1. Concept Design Process

Dashboard views

View	Name	Description
1	**Market Overview**	Comparison between 2 to 18 coins from 5 different market analysis angles.
2	**Social Impact Analyzer**	In-depth analysis of key social metrics and interactive study of correlation between price and social talk volume, per coin and month.
3	**Perspectivizer**	Tool that helps to estimate a Token A price if it had the Market Capitalization of a Token B.
4	**Tweets Language Processing**	Results of viral analysis and Natural Language Processing, both own-implemented, of influencers tweets.

1. Concept Design Process

Market Overview: *Datasets 1 & 2*

- Sentiment Overview, Market Cap, Developers Activity (Show **3-month aggregate**)
- Financial Analysis, On-Chain Analysis (Show **time-series with daily granularity**)

Possibility of selecting **multiple coins** to be compared in all 5 modules.

9X1620				SANTIMENT					LUNAR CRUSH / SANTIMENT
TOTAL PER DAY	Asset	Month	Day	*Sentiment Positive*	*Sentiment Negative*	*Developer Activity*	*Dev. activ. change 30d*	*Contributors count*	*Market Cap*
	BTC	1	1						
	BTC	1	2	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...
	BTC	3	30	...	...	...	...	...	...
	ETH	1	1	...	...	...	...	...	...
	...	...	...						

11X1620				SANTIMENT				LUNAR CRUSH			
TOTAL PER DAY	Asset	Month	Day	Active Addresses 1 day	Exchange balance	Network Growth	Transaction Volume	Perpetuals funding rate	Price	Market Cap	Volume 24h
	BTC	1	1	...	...	...	...	...	...	...	...
	BTC	1	2	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...
	BTC	3	30	...	...	...	...	...	...	...	...
	ETH	1	1	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...

1. Concept Design Process

Market Overview: *Datasets 1 & 2*

- Sentiment Overview, Market Cap, Developers Activity (Show **3-month aggregate**)
- Financial Analysis, On-Chain Analysis (Show **time-series with daily granularity**)

Possibility of selecting **multiple coins** to be compared in all 5 modules.

9X1620				SANTIMENT					LUNAR CRUSH / SANTIMENT
TOTAL PER DAY	Asset	Month	Day	*Sentiment Positive*	*Sentiment Negative*	*Developer Activity*	*Dev. activ. change 30d*	*Contributors count*	*Market Cap*
	BTC	1	1						
	BTC	1	2	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...
	BTC	3	30	...	...	...	...	...	...
	ETH	1	1	...	...	...	...	...	...
	...	...	...						

11X1620				SANTIMENT					LUNAR CRUSH		
TOTAL PER DAY	Asset	Month	Day	Active Addresses 1 day	Exchange balance	Velocity	Transaction Volume	Perpetuals funding rate	Price	Market Cap	Volume 24h
	BTC	1	1	...	...	...	...	...	...	...	...
	BTC	1	2	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...
	BTC	3	30	...	...	...	...	...	...	...	...
	ETH	1	1	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...

1. Concept Design Process

Social Impact: *Dataset 3*

- Social Scope Metrics (Show **4 scoreboards** with **3-month aggregate**)
- Price VS Social Volume (Show **time-series with daily granularity**)

Possibility of selecting **multiple coins** and **months** to be compared in both modules.

16X1620				SANTIMENT				LUNAR CRUSH								
TOTAL PER DAY	Asset	Month	Day	Price	Bitcointalk Volume	Telegram Volume	Reddit Volume	Twitter Volume	Youtube Volume	News Volume	Medium Volume	Market Dominance	Social Dominance	Average Sentiment	Correlation Rank	Shared Links
	BTC	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...
	BTC	1	2	...	...	...	...	...	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	BTC	3	30	...	...	...	...	...	...	...	...	...	...	...	...	...
	ETH	1	1	...	...	...	...	...	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

1. Concept Design Process

Perspectivizer: *Dataset 4*

- Social Scope Metrics (Show **4 scoreboards** with **3-month aggregate**)
- Price VS Social Volume (Show **time-series with daily granularity**)

Possibility of selecting **multiple coins** and **months** to be compared in both modules.

DASHBOARD 3				
4X18		LUNAR CRUSH / SANTIMENT		SANTIMENT
DAY DECEMBER 1	Asset	Price	Market Cap	Circulating supply
		1 DAY ABSOLUTE	*1 DAY ABSOLUTE*	*1 DAY ABSOLUTE*
	BTC	...	...	...
	ETH	...	...	...
	ADA	...	...	...

1. Concept Design Process

Perspectivizer: *Dataset 4*

- Social Scope Metrics (Show **4 scoreboards** with **3-month aggregate**)
- Price VS Social Volume (Show **time-series with daily granularity**)

Possibility of selecting **multiple coins** and **months** to be compared in both modules.

DASHBOARD 3				
4X18		LUNAR CRUSH / SANTIMENT		SANTIMENT
DAY DECEMBER 1	Asset	Price	Market Cap	Circulating supply
		1 DAY ABSOLUTE	*1 DAY ABSOLUTE*	*1 DAY ABSOLUTE*
	BTC	...	...	...
	ETH	...	...	...
	ADA	...	...	...

1. Concept Design Process

Tweets Language Processing: *Datasets 5,6,7*

- **Cloud Word** with most popular terms.
- **Top 5 Influencers** and sentiment analysis distribution of their tweets.
- **Top 5 Tweets** by own-computed Viral Score.

Possibility of selecting **multiple coins** to be compared in both modules.

COIN	INFLUENCER	POSITIVE	NEGATIVE	NEUTRAL
btc	a			
btc	b			
btc	c			
btc	d			
btc	e			
eth	f			
eth	g			
eth	h			
eth	i			
eth	j			

COIN	RANK	TWEET	(1*RTs+0.5*FAVS) METRIC	SENTIMENT
btc	1	blabla	90432	Positive
btc	2	blabla	69420	Neutral
btc	3	blabla	20400	Negative
btc	4	blabla	7888	Neutral
btc	5	blabla	420	Positive
eth	1	blabla	90432	Neutral
eth	2	blabla	69420	Negative
eth	3	blabla	20400	Positive
eth	4	blabla	7888	Neutral
eth	5	blabla	420	Negative

15 top words

WORDCLOUD

COIN	WORD	COUNT
btc	a	222
btc	b	111
btc	c	69
btc	d	42
btc		
btc		
btc		
eth	a	222
eth	b	111
eth	c	69
eth	d	42
eth		
eth		

Google Data Studio Dashboard

2. Google Data Studio Dashboard

Possible approaches

Code it with python libraries:

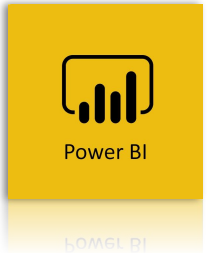
- Classical (matplotlib, seaborn…)
- Dash
- Streamlit

2. Google Data Studio Dashboard

Possible approaches

Code it with python libraries:

- Classical (matplotlib, seaborn…)
- Dash
- Streamlit



Final result

https://datastudio.google.com/s/qEaOm08vdWs

Scraping strategy and implementation

3. Scraping strategy and implementation

Sources

COIN OVERVIEW *(Dashboard 1)*		SOCIAL IMPACT *(Dashboard 2)*		PERSPECTIVIZER *(Dashboard 3)*	NLP ANALYSIS *(Dashboard 4)*	
DEV. METRICS	ON-CHAIN METRICS	SOCIAL VOLUME	SOCIAL METRICS	MARKET METRICS	TWEETS	TOP INFLUENCERS
·santiment·	kraken g	·santiment·	LunarCrush	·santiment·		LunarCrush

3. Scraping strategy and implementation

Implementation

Several APIs have been considered throughout the project and their respective wrappers have been implemented. Some have been improved to minimize the computational time and to avoid rate limit.

Custom API Wrappers

(*) santiment

(!!) LunarCrush

(**) g

(**) kraken

API Wrappers

(*) Tweepy

(**)

Web Scraping

twint (!)

(*) Insufficient recent data (**) Insufficient public data
(!) Inefficient (!!) Server errors

Natural Language Processing

4. Natural Language Processing

The data flow

Who?

Top 10 influencers from each coin

What?

Tweets where they talk about their coin

Why?

Sentiment Analysis and Cloud Word

4. Natural Language Processing

The Data Flow

Clean tweets

(#, @, $, urls, :-), !?, ...)

Sentiment Analysis

Word Cloud

spaCy

4. Natural Language Processing

The Sentiment Analysis

Hugging Face is a startup in the Natural Language Processing domain that offers open-source libraries and pre-trained models for the general public.

Which resources do we use from it?

- ***Transformers***: library that provides thousands of pre-trained models and a pipeline to perform tasks on different areas such as text, vision, and audio.

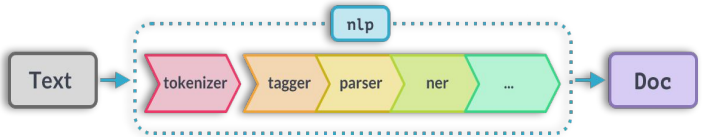
- ***Beto-sentiment-analysis***: sentiment analysis pre-trained model part of *pysentimiento, a* Python toolkit for Sentiment Analysis and Social NLP tasks.

4. Natural Language Processing

The Word Cloud

SpaCy is a free open-source library for Natural Language Processing in Python that features word counts for building word clouds.

SpaCy NLP() built-in pipeline



Specific stopwords removal

Iterative process to remove **non-insightful words**, such as: coin names, currency, crypto, usd, ...

Model Selection

en_core_web_sm

Extracting popular terms

Rank words by count

Recap and potential improvements opportunities

5. Recap and potential improvements opportunities

Recap

We implemented...

4 custom APIs wrappers and used **2** APIs with a total of **39** different endpoints,
1 asynchronous parallel web scraper.

We scraped ...

More than **50** different cryptocurrency metrics for **+90 days** and **+30,000 tweets** stored

We implemented …

1 NLP pipeline with **8** different analysis steps + **1** Sentiment Analysis model transformer

We designed ...

4 dashboard views with **3** tables, **14** scoreboards, **3** time-series, **1** barchart, **1** rectangles plot, **1** word cloud, **6** filter controls and **2** community visualizations.

5. Recap and potential improvements opportunities

Improvement Opportunities

Some coins were missing specific metric values.

Scrape more tweets over a longer period of time.

Further research in NLP.

Connect dashboard to online database that gets updated every day.

Scoreboards showing coin name and logo, dynamically depending on the asset selected.

Other state-of-the-art solutions

6. Other state-of-the-art solutions

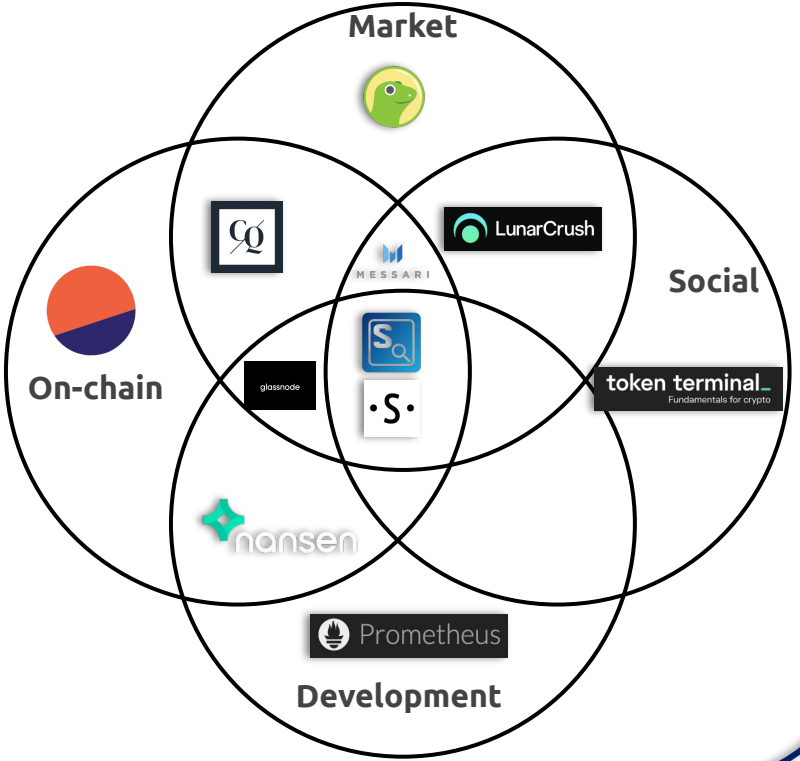
Market Overview and **Social Impact Analyzer**

We are better in …

- Simplified, user-friendly interface.
- Aggregation of multiple angles.
- Insightful overview.
- Compare different social volumes in same visualization.

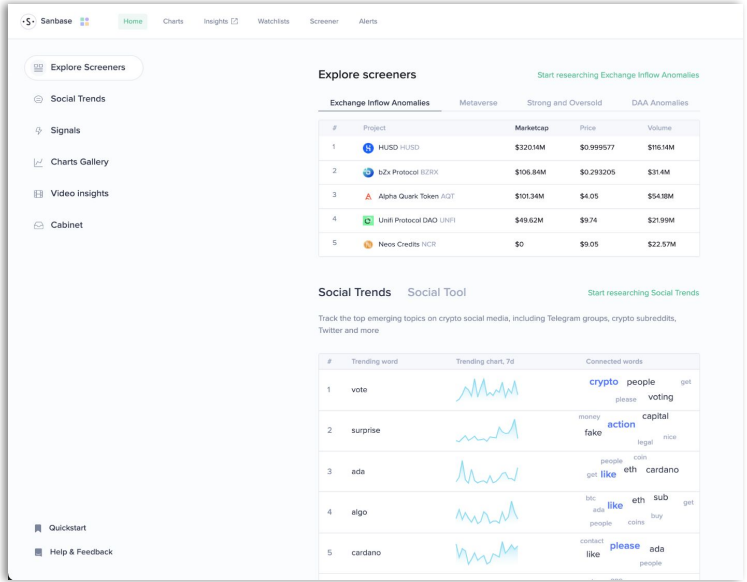
We lack …

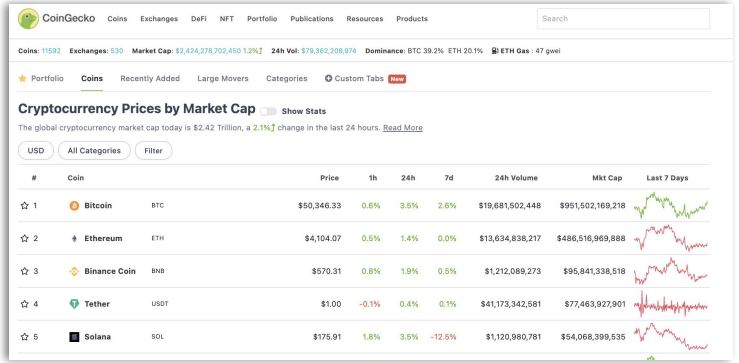
- We only show 18 coins with some metrics.
- Dashboard loading times.
- Presence of missing values.



6. Other state-of-the-art solutions

Market Overview and Social Impact Analyzer

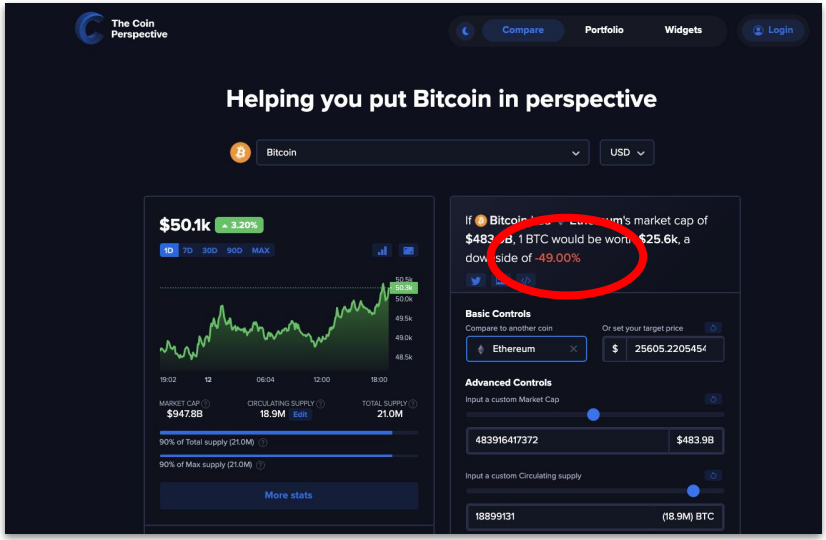




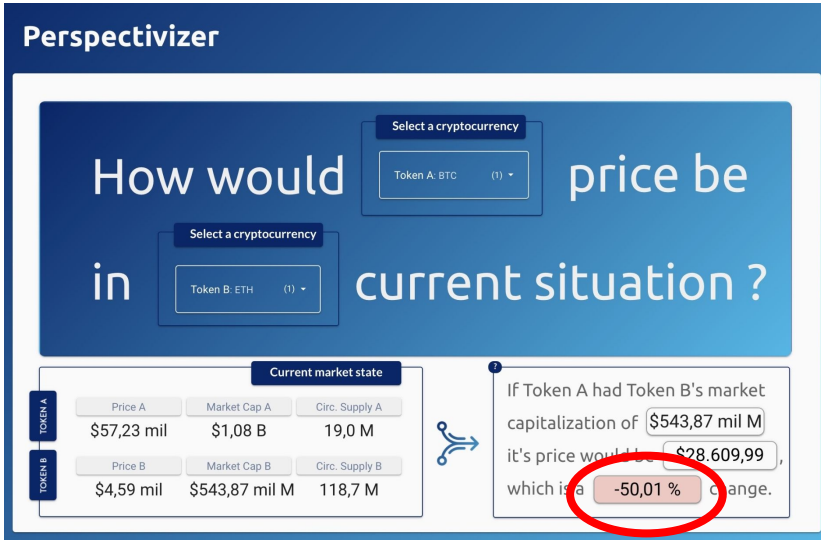
6. Other state-of-the-art solutions

Perspectivizer

Go-to solution in the market:



Our solution:



Attributions and references

7. Attributions and references

Icons

Flaticon repository

Data Scraping

- tweepy: Twitter for Python!
- twint: Asynchronus Twitter scraping
- sanpy: Santiment API Python Client
- pycoingecko: Python wrapper for the CoinGecko API

NLP

- pysentimiento: A Python Toolkit for Sentiment Analysis and Social NLP tasks
- HugginFace: State-of-the-art NLP

THANK YOU