Sprint Review 2

Thursday, 1st October

Hello Again 👋

Jamie

Quick Outcomes 🙌

- Clearly defined MVP
- Overall UI starting to take shape
- Backend libraries delivering usable data
- Decided on working product architecture
- Bonded as a team

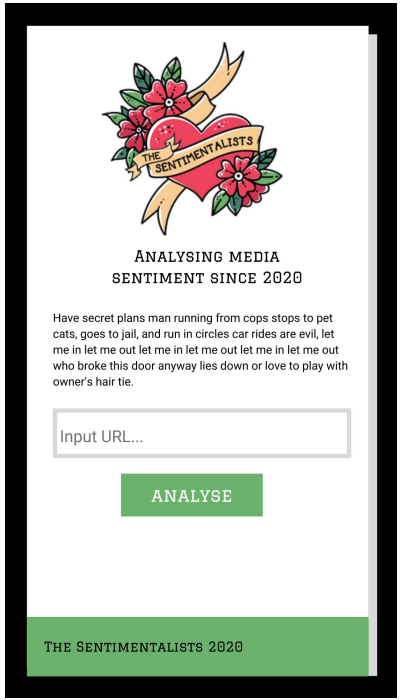
Jamie

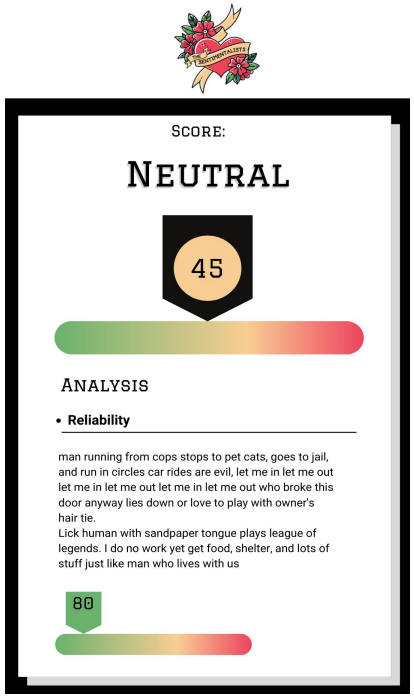
Frontend 👩‍💻

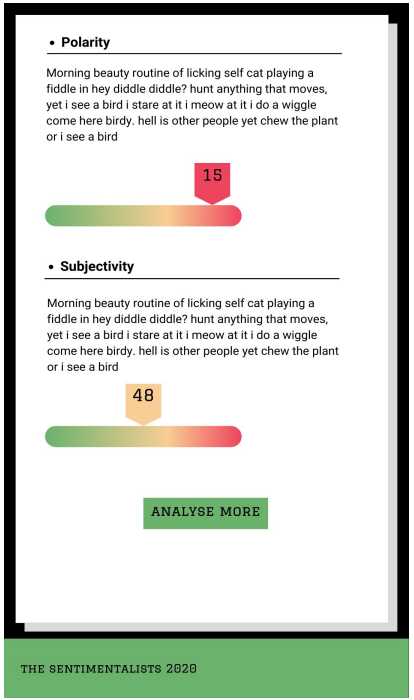
- Developed UI prototype with colour
- Defined the components needed to display data
- Decided app will be built in React using Material for styling and layout, with additional CSS where appropriate
- Started implementing grid layout for app
- Gained a greater understanding of props, state and event handling and where they fit into our product functionality

Ellie

User Interface - Mobile

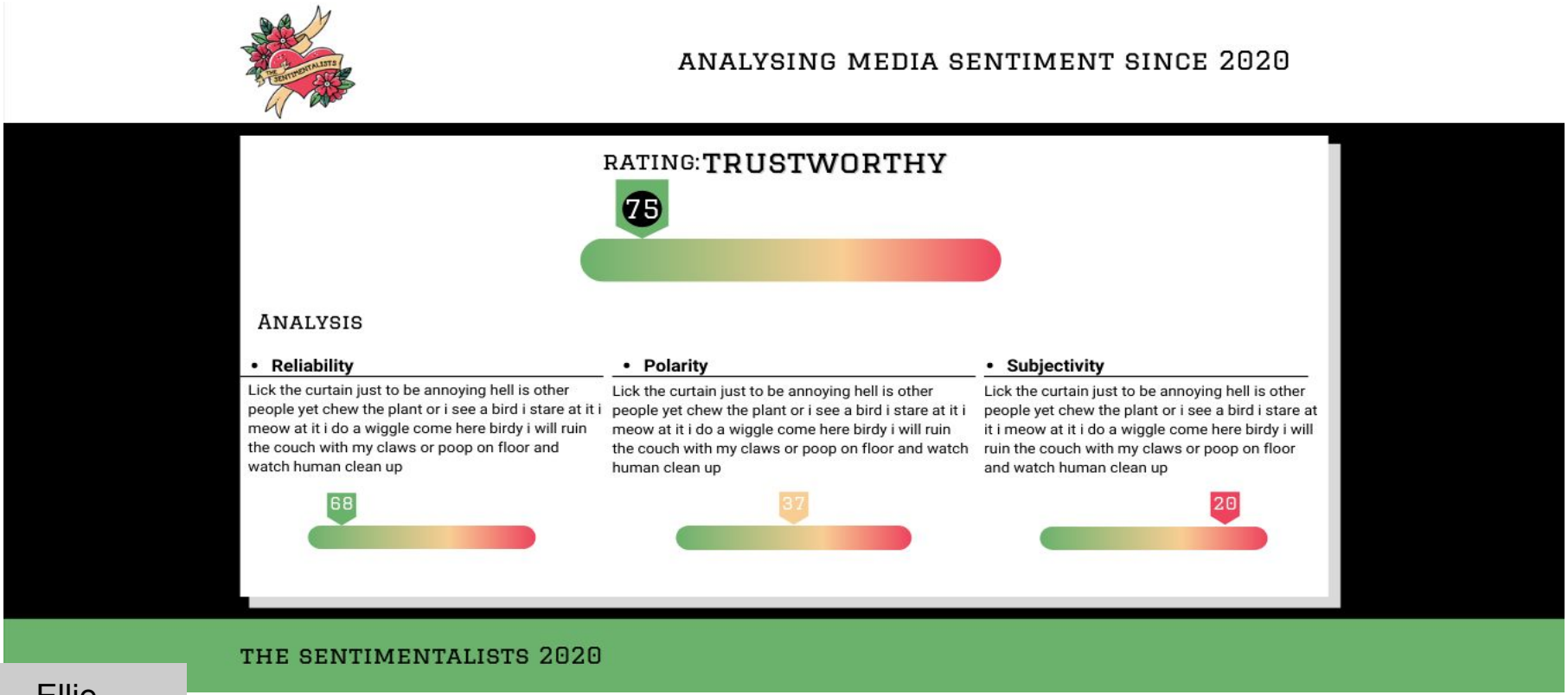






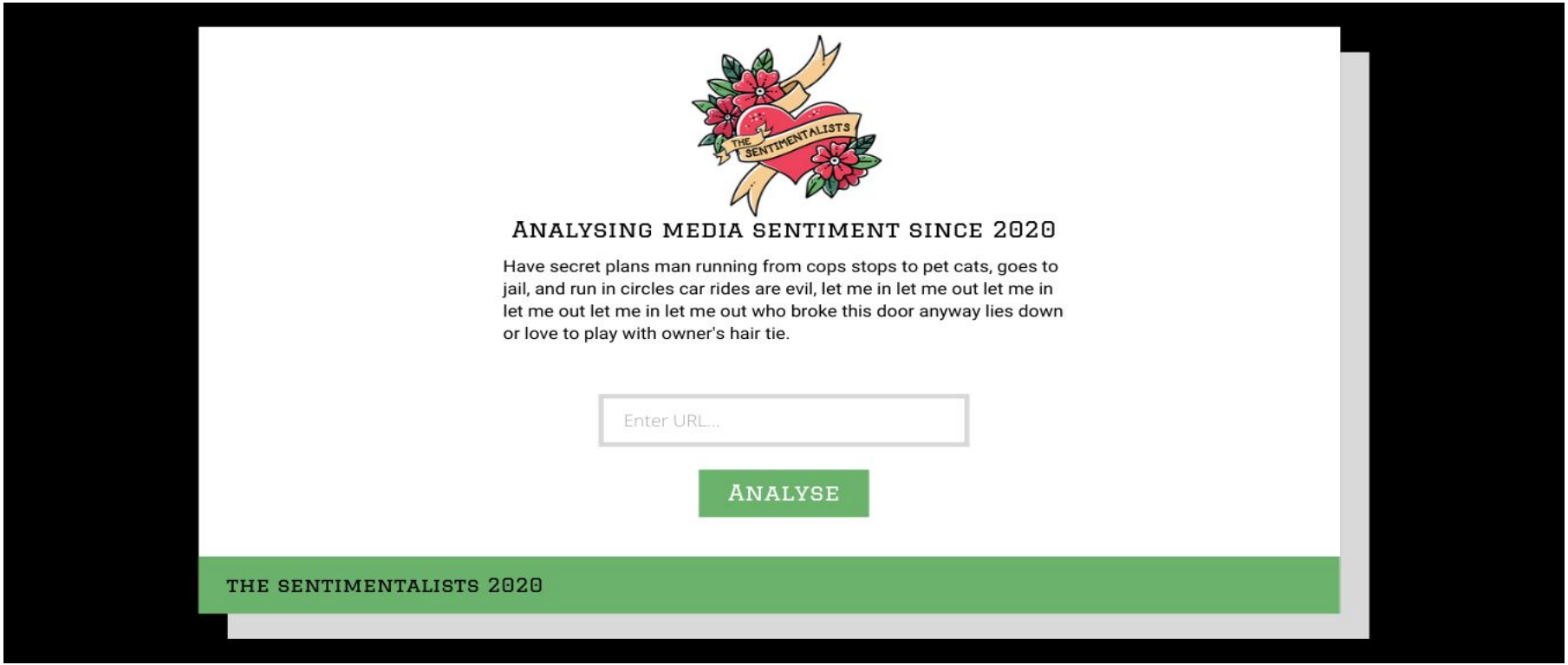
Ellie

User Interface - Large Screens



Ellie

User Interface - Large Screens



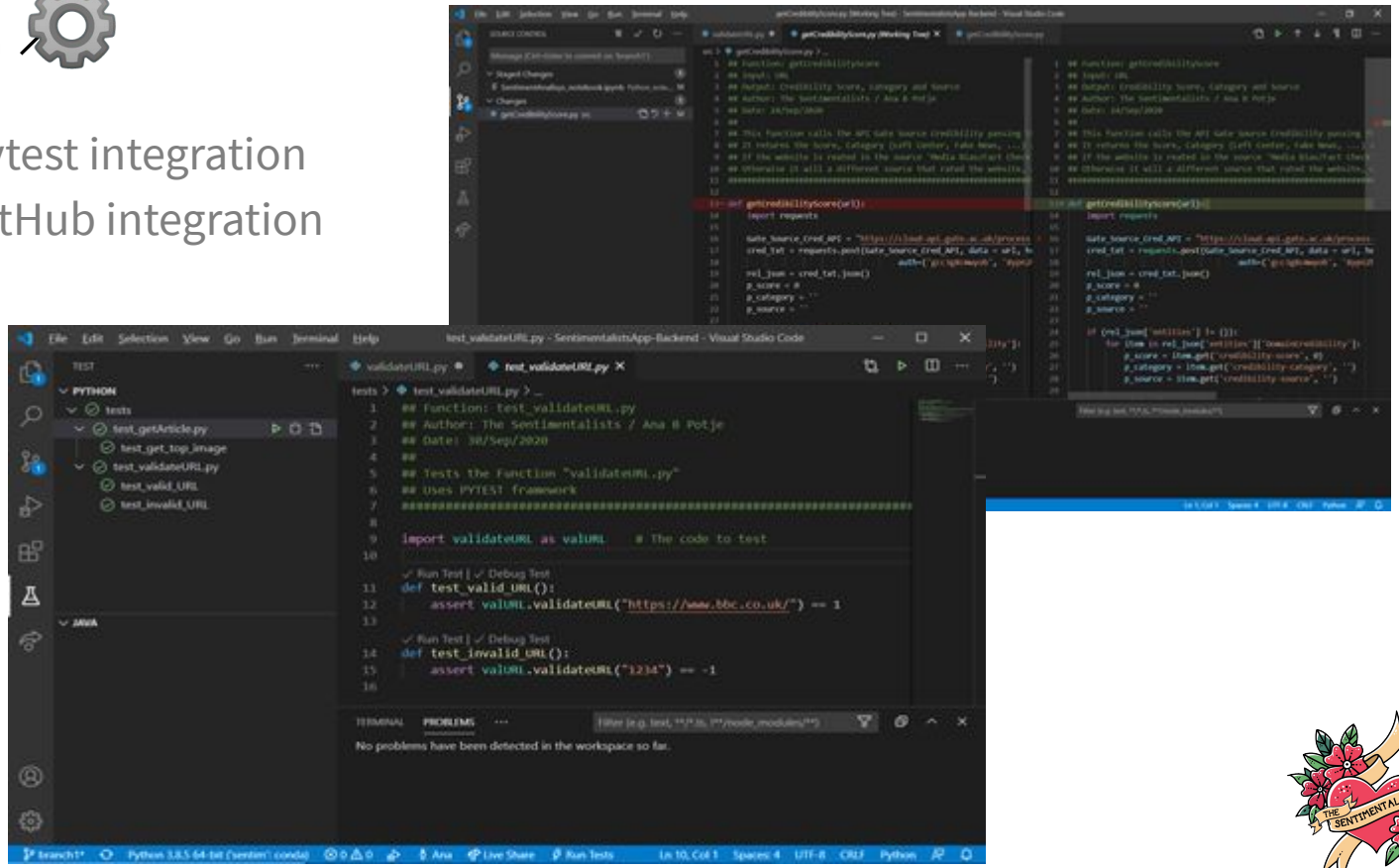
Ellie

Backend ⚙️

- Input for modules will be an **URL** (no free text / document upload at this stage)
- Defined **Python** as backend programming language
- Tools and Frameworks:
  - Visual Studio Code;
  - Visual Studio Live Share (collaboration);
  - Jupyter Notebook;
  - PyTest (testing);
  - GitHub (repositories and version control);
  - GitHub Project Board (track project tasks);

Backend

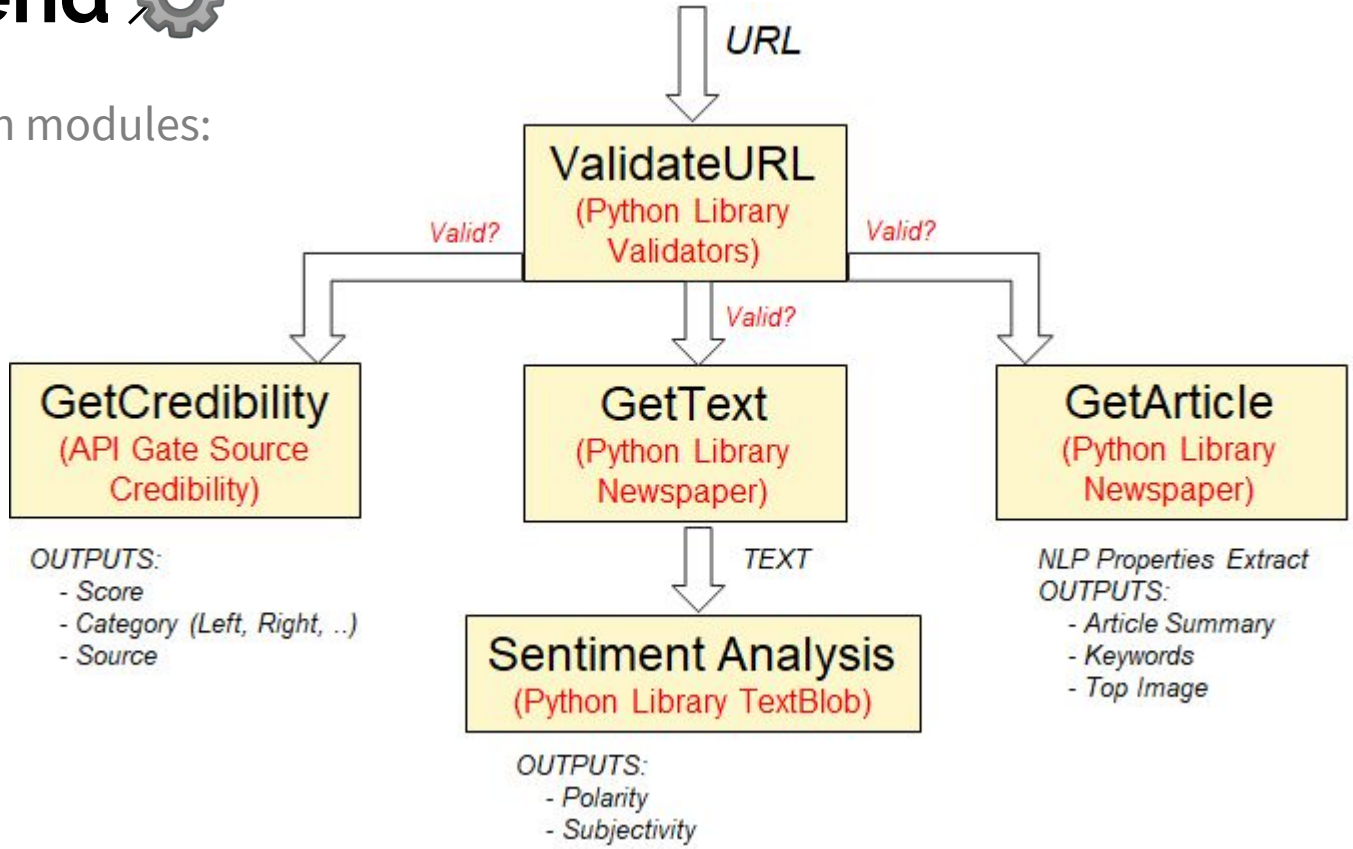
- VsCode X Pytest integration
- VsCode X GitHub integration



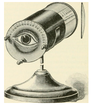
Ana

Backend

- Python modules:

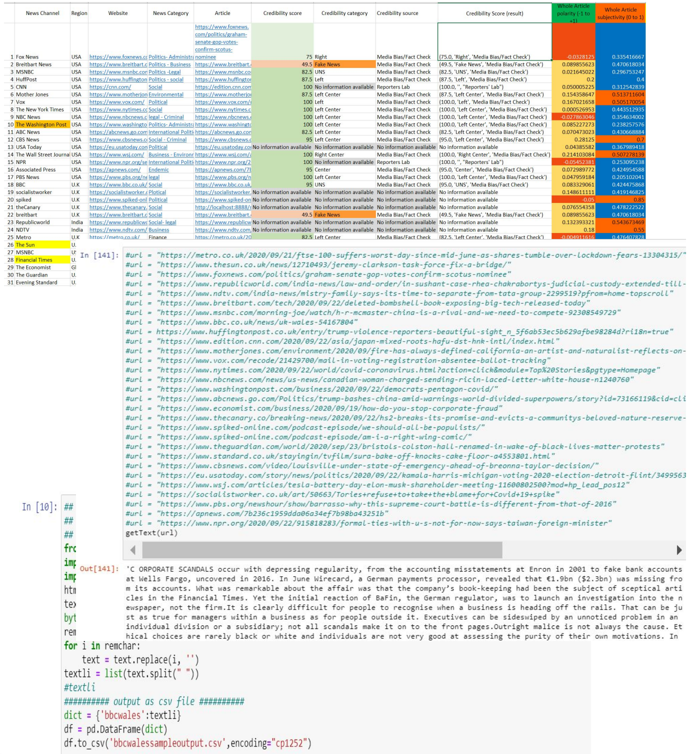


Backend



Tests, tests, tests!

- **Tested Sentiment Analysis function** - noted Polarity, Sensitivity results from 30 News articles with varying sources, subjects
- **Tested Credibility Score function** - credibility scores, category and source results noted
- Discussed **results and observations** with team to draw **further actions** (validate URL, error causing article, follow-up with Gate Cloud solution)
- **Created** Python code removing unwanted characters from text to display



	News Channel	Region	Website	News Category	Article	Credibility score	Credibility category	Credibility source	Credibility Score (result)	Whole Article polarity (-1 to +1)	Whole Article subjectivity (0 to 1)
1	Fox News	USA	https://www.foxnews.cc	Politics- Administra	https://www.foxnews.com/politics/graham-senate-gop-votes-confirm-scotus-nominee	75	Right	Media Bias/Fact Check	(75.0, 'Right', 'Media Bias/Fact Check')	-0.0328125	0.335416667
2	Breitbart News	USA	https://www.breitbart.c	Politics - Business	https://www.breitbart.	49.5	Fake News	Media Bias/Fact Check	(49.5, 'Fake News', 'Media Bias/Fact Check')	0.089855623	0.470618034
3	MSNBC	USA	https://www.msnbc.con	Politics -Legal	https://www.msnbc.co	82.5	UNS	Media Bias/Fact Check	(82.5, 'UNS', 'Media Bias/Fact Check')	0.021645022	0.296753247
4	HuffPost	USA	https://www.huffington	Politics - social	https://www.huffingtor	87.5	Left	Media Bias/Fact Check	(87.5, 'Left', 'Media Bias/Fact Check')	0.2	0.4
5	CNN	USA	https://cnn.com/	Social	https://edition.cnn.com	100	No information available	Reporters Lab	(100.0, '', "Reporters' Lab")	0.050005225	0.312542839
6	Mother Jones	USA	https://www.motherjon	Environmental	https://www.motherjo	87.5	Left Center	Media Bias/Fact Check	(87.5, 'Left Center', 'Media Bias/Fact Check')	0.154358647	0.513711604
7	Vox	USA	https://www.vox.com/	Political	https://www.vox.com/	100	Left	Media Bias/Fact Check	(100.0, 'Left', 'Media Bias/Fact Check')	0.167021658	0.505170054
8	The New York Times	USA	https://www.nytimes.cc	Social	https://www.nytimes.c	100	Left Center	Media Bias/Fact Check	(100.0, 'Left Center', 'Media Bias/Fact Check')	0.000526953	0.443512935
9	NBC News	USA	https://www.nbcnews.c	legal - Criminal	https://www.nbcnews.	100	Left Center	Media Bias/Fact Check	(100.0, 'Left Center', 'Media Bias/Fact Check')	-0.027863046	0.354634002
10	The Washington Post	USA	https://www.washingto	Politics- Administra	https://www.washingto	100	Left Center	Media Bias/Fact Check	(100.0, 'Left Center', 'Media Bias/Fact Check')	0.085227273	0.238257576
11	ABC News	USA	https://abcnews.go.com	International Politi	https://abcnews.go.cor	82.5	Left Center	Media Bias/Fact Check	(82.5, 'Left Center', 'Media Bias/Fact Check')	0.070473023	0.430668884
12	CBS News	USA	https://www.cbsnews.c	Social - Criminal	https://www.cbsnews.	95	Left Center	Media Bias/Fact Check	(95.0, 'Left Center', 'Media Bias/Fact Check')	0.28125	0.7
13	USA Today	USA	https://eu.usatoday.con	Political	https://eu.usatoday.co	No information available	No information available	No information available	No information available	0.04385582	0.367989418
14	The Wall Street Journal	USA	https://www.wsj.com/	Business - Environr	https://www.wsj.com/	100	Right Center	Media Bias/Fact Check	(100.0, 'Right Center', 'Media Bias/Fact Check')	0.214103084	0.507278139
15	NPR	USA	https://www.npr.org/se	International Politi	https://www.npr.org/2	100	No information available	Reporters Lab	(100.0, '', "Reporters' Lab")	-0.05452381	0.253095238
16	Associated Press	USA	https://apnews.com/	Endemic	https://apnews.com/7	95	Center	Media Bias/Fact Check	(95.0, 'Center', 'Media Bias/Fact Check')	0.072989772	0.424954588
17	PBS News	USA	https://www.pbs.org/ne	legal	https://www.pbs.org/n	100	Left Center	Media Bias/Fact Check	(100.0, 'Left Center', 'Media Bias/Fact Check')	0.047959184	0.205102041
18	BBC	U.K	https://www.bbc.co.uk/	Social	https://www.bbc.co.uk	95	UNS	Media Bias/Fact Check	(95.0, 'UNS', 'Media Bias/Fact Check')	0.083329061	0.424475868
19	socialistworker	U.K	https://socialistworker.	Plotical	https://socialistworker.	No information available	No information available	No information available	No information available	0.148611111	0.419146825
20	spiked	U.K	https://www.spiked-onl	Political	https://www.spiked-on	No information available	No information available	No information available	No information available	-0.05	0.85
21	theCanary	U.K	https://www.thecanary.	Social	http://localhost:8888/	No information available	No information available	No information available	No information available	0.076554358	0.478222522
22	breitbart	U.K	https://www.breitbart.c	Social	https://www.breitbart.	49.5	Fake News	Media Bias/Fact Check	(49.5, 'Fake News', 'Media Bias/Fact Check')	0.089855623	0.470618034
23	Republicworld	India	https://www.republicwc	Social- legal	https://www.republicw	No information available	No information available	No information available	No information available	0.132393321	0.543673469
24	NDTV	India	https://www.ndtv.com/	Business	https://www.ndtv.com	No information available	No information available	No information available	No information available	0.18	0.55
25	Metro	U.K	https://metro.co.uk/	Finance	https://metro.co.uk/20	82.5	Left Center	Media Bias/Fact Check	(82.5, 'Left Center', 'Media Bias/Fact Check')	-0.004911616	0.476407828
26	The Sun	U.									
27	MSNBC	U:									
28	Financial Times	U.									
29	The Economist	Gl									
30	The Guardian	U.									
31	Evening Standard	U.									

In [141]:

```
#url = "https://metro.co.uk/2020/09/21/ftse-100-suffers-worst-day-since-mid-june-as-shares-tumble-over-lockdown-fears-13304315/"
#url = "https://www.thesun.co.uk/news/12710493/jeremy-clarkson-task-force-fix-a-bridge/"
#url = "https://www.foxnews.com/politics/graham-senate-gop-votes-confirm-scotus-nominee"
#url = "https://www.republicworld.com/india-news/law-and-order/in-sushant-case-rhea-chakrabortys-judicial-custody-extended-till-
#url = "https://www.ndtv.com/india-news/mistry-family-says-its-time-to-separate-from-tata-group-2299519?pfrom=home-topscroll"
#url = "https://www.breitbart.com/tech/2020/09/22/deleted-bombshell-book-exposing-big-tech-released-today"
#url = "https://www.msnbc.com/morning-joe/watch/h-r-mcmaster-china-is-a-rival-and-we-need-to-compete-92308549729"
#url = "https://www.bbc.co.uk/news/uk-wales-54167804"
#url = https://www.huffingtonpost.co.uk/entry/trump-violence-reporters-beautiful-sight_n_5f6ab53ec5b629afbe98284d?ri18n=true"
#url = "https://www.edition.cnn.com/2020/09/22/asia/japan-mixed-roots-hafu-dst-hnk-intl/index.html"
#url = "https://www.motherjones.com/environment/2020/09/fire-has-always-defined-california-an-artist-and-naturalist-reflects-on-
#url = "https://www.vox.com/recode/21429700/mail-in-voting-registration-absentee-ballot-tracking"
#url = "https://www.nytimes.com/2020/09/22/world/covid-coronavirus.html?action=click&module=Top%20Stories&pgtype=Homepage"
#url = "https://www.nbcnews.com/news/us-news/canadian-woman-charged-sending-ricin-laced-letter-white-house-n1240760"
#url = "https://www.washingtonpost.com/business/2020/09/22/democrats-pentagon-covid/"
#url = "https://www.abcnews.go.com/Politics/trump-bashes-china-amid-warnings-world-divided-superpowers/story?id=73166119&cid=cli
#url = "https://www.economist.com/business/2020/09/19/how-do-you-stop-corporate-fraud"
#url = "https://www.thecanary.co/breaking-news/2020/09/22/hs2-breaks-its-promise-and-evicts-a-communitys-beloved-nature-reserve-
#url = "https://www.spiked-online.com/podcast-episode/we-should-all-be-populists/"
#url = "https://www.spiked-online.com/podcast-episode/am-i-a-right-wing-comic/"
#url = "https://www.theguardian.com/world/2020/sep/23/bristols-colston-hall-renamed-in-wake-of-black-lives-matter-protests"
#url = "https://www.standard.co.uk/stayingin/tvfilm/sura-bake-off-knocks-cake-floor-a4553801.html"
#url = "https://www.cbsnews.com/video/louisville-under-state-of-emergency-ahead-of-breonna-taylor-decision/"
#url = "https://eu.usatoday.com/story/news/politics/2020/09/22/kamala-harris-michigan-voting-2020-election-detroit-flint/3499563
#url = "https://www.wsj.com/articles/tesla-battery-day-elon-musk-shareholder-meeting-11600802500?mod=hp_lead_pos12"
#url = "https://socialistworker.co.uk/art/50663/Tories+refuse+to+take+the+blame+for+Covid+19+spike"
#url = "https://www.pbs.org/newshour/show/barrasso-why-this-supreme-court-battle-is-different-from-that-of-2016"
#url = "https://apnews.com/7b236c1959dda06a34ef7b98ba43251b"
#url = "https://www.npr.org/2020/09/22/915818283/formal-ties-with-u-s-not-for-now-says-taiwan-foreign-minister"
getText(url)
```

Out[141]: 'C ORPORATE SCANDALS occur with depressing regularity, from the accounting misstatements at Enron in 2001 to fake bank accounts at Wells Fargo, uncovered in 2016. In June Wirecard, a German payments processor, revealed that €1.9bn ($2.3bn) was missing fro m its accounts. What was remarkable about the affair was that the company's book-keeping had been the subject of sceptical arti cles in the Financial Times. Yet the initial reaction of BaFin, the German regulator, was to launch an investigation into the n ewspaper, not the firm.It is clearly difficult for people to recognise when a business is heading off the rails. That can be ju st as true for managers within a business as for people outside it. Executives can be sideswiped by an unnoticed problem in an individual division or a subsidiary; not all scandals make it on to the front pages.Outright malice is not always the cause. Et hical choices are rarely black or white and individuals are not very good at assessing the purity of their own motivations. In

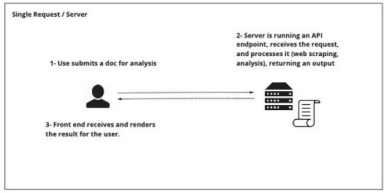
In [10]:

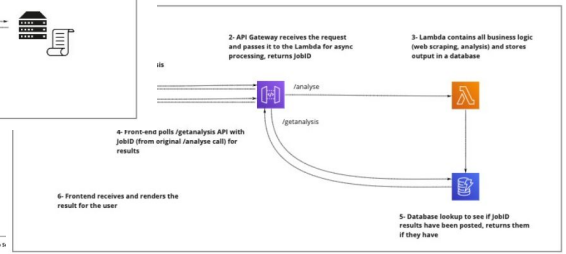
```
##
##
##
fro
imp
imp
htm
tex
byt
rem
for i in remchar:
    text = text.replace(i, '')
textli = list(text.split(" "))
#textli
########## output as csv file ##########
dict = {'bbcwales':textli}
df = pd.DataFrame(dict)
df.to_csv('bbcwalessampleoutput.csv',encoding="cp1252")
```

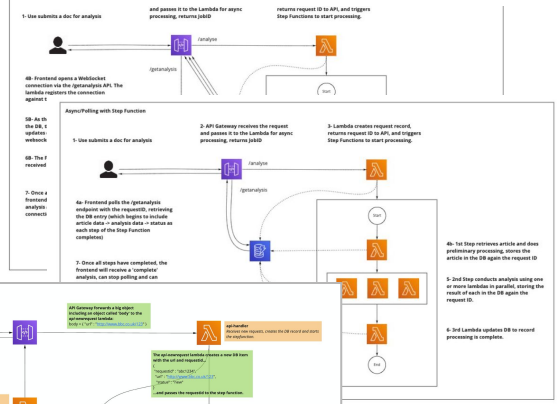
Jyothsna

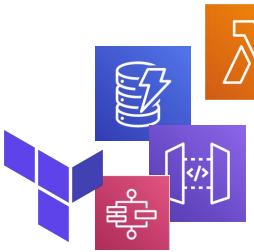
Infrastructure

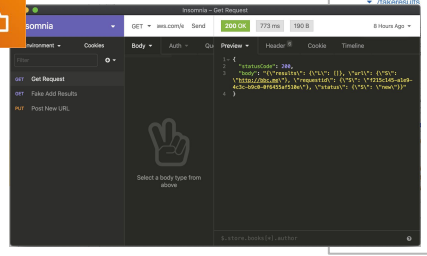
- Explored different architecture patterns
- Reviewed pros and cons of each for our needs
- Mocked up a couple of the designs in AWS
- Implementing the final architecture using AWS API Gateway, Lambdas (Python), DynamoDB (NoSQL) and Step Functions, deployed with Terraform

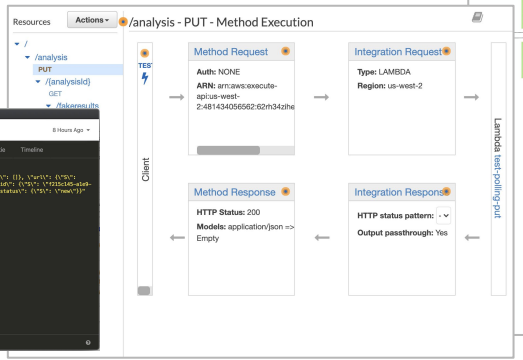


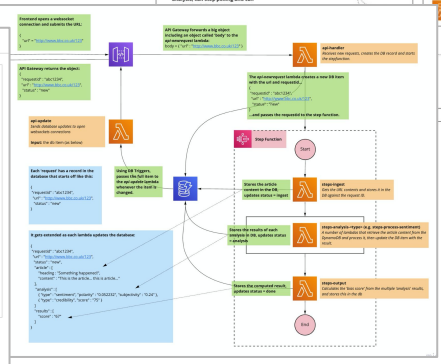








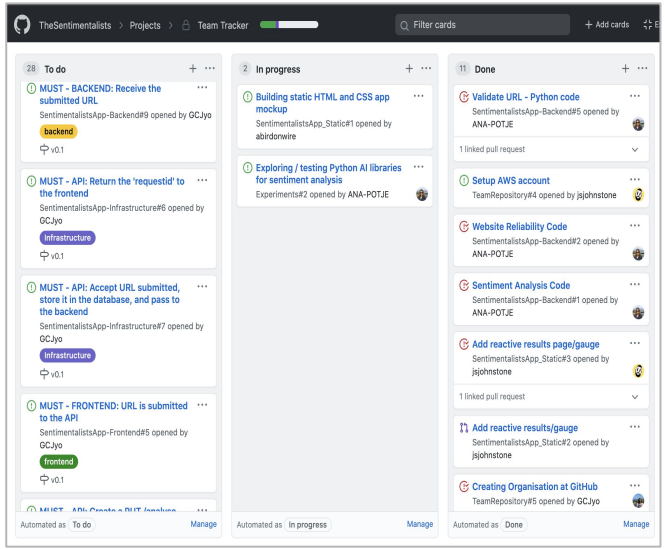


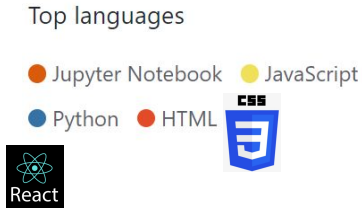


Jamie

Team Practices

- Daily standups, formal retro meetings, test driven development, Pair programming as we move into V1.0
- **Github Team tracker** to priorities tasks in each sprint
- Collaboration: Reviewed various ways to **display results** at frontend and agreed one particular method to present the information



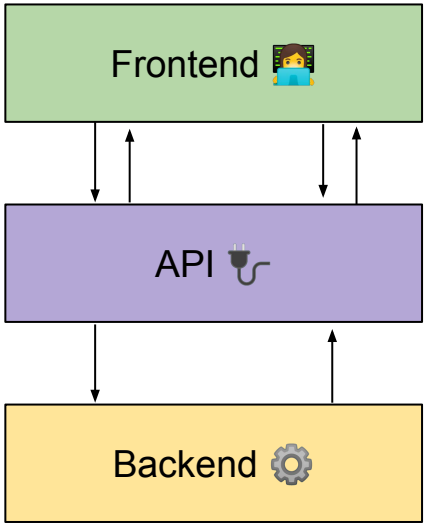


A First Draft 🎯

Agreed our first milestone/set of features, getting structure in place and our first data flowing:

Version 0.1:

- A **simple front-end** allowing the user to enter and **submit a URL** and **view analysis results** (just ‘website credibility’ to get us started).
- An API layer providing **two endpoints**: one to submit a URL/the request, and another to poll for results
- A backend that retrieves **the credibility score** for the given URL and **stores results** against the request



Ana

The First Major Milestone 🌟

Extending functionalities to create a useful tool

Version 1.0

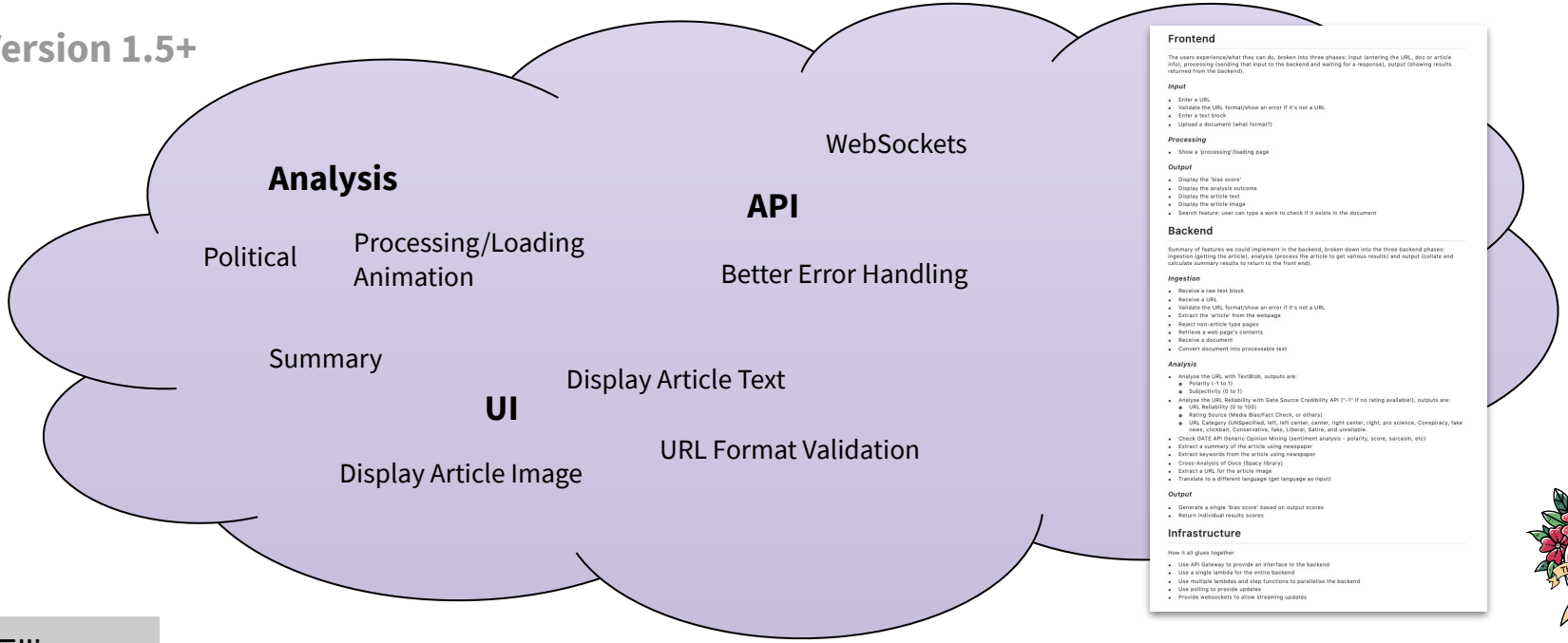
- Retrieve and store the article content
- Add additional analysis, such as polarity, subjectivity
- Calculate a cumulative ‘bias score’

Jyothsna

The Future Goals 🚀

Many other features explored which we will consider for future iterations

Version 1.5+



Frontend

The users experience/what they can do, broken into three phases: input (entering the URL, doc or article info), processing (sending that input to the backend and waiting for a response), output (showing results returned from the backend).

Input

- Enter a URL
- Validate the URL format/show an error if it's not a URL
- Enter a text block
- Upload a document (what format?)

Processing

- Show a 'processing'/loading page

Output

- Display the 'bias score'
- Display the analysis outcome
- Display the article text
- Display the article image
- Search feature: user can type a work to check if it exists in the document

Backend

Summary of features we could implement in the backend, broken down into the three backend phases: ingestion (getting the article), analysis (process the article to get various results) and output (collate and calculate summary results to return to the front end).

Ingestion

- Receive a raw text block
- Receive a URL
- Validate the URL format/show an error if it's not a URL
- Extract the 'article' from the webpage
- Reject non-article type pages
- Retrieve a web page's contents
- Receive a document
- Convert document into processable text

Analysis

- Analyse the URL with TextBlob, outputs are:
  - Polarity (-1 to 1)
  - Subjectivity (0 to 1)
- Analyse the URL Reliability with Gate Source Credibility API ("-1" if no rating available!), outputs are:
  - URL Reliability (0 to 100)
  - Rating Source (Media Bias/Fact Check, or others)
  - URL Category (UNSpecified, left, left center, center, right center, right, pro science, Conspiracy, fake news, clickbait, Conservative, fake, Liberal, Satire, and unreliable.
- Check GATE API Generic Opinion Mining (sentiment analysis - polarity, score, sarcasm, etc)
- Extract a summary of the article using newspaper
- Extract keywords from the article using newspaper
- Cross-Analysis of Docs (Spacy library)
- Extract a URL for the article image
- Translate to a different language (get language as input)

Output

- Generate a single 'bias score' based on output scores
- Return individual results scores

Infrastructure

How it all glues together

- Use API Gateway to provide an interface to the backend
- Use a single lambda for the entire backend
- Use multiple lambdas and step functions to parallelise the backend
- Use polling to provide updates
- Provide websockets to allow streaming updates

Ellie

Next Sprint Goals 🏆

- Get through the v0.1 tickets

Jamie

Questions?