Software Development Report

Francesco, Fabio, Se\_an

February 2019

1

Contents

1 Introduction 5

1.1 Aim . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

2 Project Managment 6

3 Sprint 1 6

3.1 Group formation 12/02/19 9.00-11.00 . . . . . . . . . . . . . . . 6

3.2 Membership . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

3.3 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

3.4 Meeting 1; 14/02/19 9.00-11.00 . . . . . . . . . . . . . . . . . . . 7

3.4.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.4.2 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.4.3 Work to-do . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.5 Meeting 2; 15/ 02/19 12.00-18.00 . . . . . . . . . . . . . . . . . . 7

3.5.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.5.2 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.5.3 Work to-do . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.6 Meting 3; 18/02/19 17.00-19.00 . . . . . . . . . . . . . . . . . . . 7

3.6.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

3.6.2 Decisions made . . . . . . . . . . . . . . . . . . . . . . . . 8

3.6.3 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.7 Meeting 4; 19/02/19 9.00-11.00 . . . . . . . . . . . . . . . . . . . 8

3.7.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.7.2 Decisions made . . . . . . . . . . . . . . . . . . . . . . . . 8

3.7.3 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.8 Meeting 4 20/02/19 13.00-13.30 . . . . . . . . . . . . . . . . . . . 8

3.8.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.8.2 To-do . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

3.8.3 Wishlist additions . . . . . . . . . . . . . . . . . . . . . . 9

3.9 Meeting 5 21/01/19 9.00-11.00 . . . . . . . . . . . . . . . . . . . 9

3.9.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

3.9.2 Decisions made . . . . . . . . . . . . . . . . . . . . . . . . 9

3.9.3 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 9

4 Sprint 2 10

4.1 Meeting 6 26/02/19 9.00-11.00 . . . . . . . . . . . . . . . . . . . 10

4.1.1 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . 10

4.1.2 Work done . . . . . . . . . . . . . . . . . . . . . . . . . . 11

4.2 Latest report . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

4.3 Meeting 7/3/19 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

5 Sprint 3 12

5.1 Meeting 16/3/19 . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2

6 Sprint 4 13

6.1 Meeting 2/4/19 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

6.2 Meeting 12/4/19 . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

7 Burndown Chart 14

8 Development Operations (Dev-ops) 16

8.1 Technologies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

8.2 virtual Environment . . . . . . . . . . . . . . . . . . . . . . . . . 16

8.3 GitHub Repository . . . . . . . . . . . . . . . . . . . . . . . . . . 16

8.3.1 Repository Structure . . . . . . . . . . . . . . . . . . . . . 17

8.4 Amazon AWS Server . . . . . . . . . . . . . . . . . . . . . . . . . 17

8.4.1 Custom Packages installed . . . . . . . . . . . . . . . . . . 18

8.5 Amazon RDS MySQL Database . . . . . . . . . . . . . . . . . . 20

8.5.1 static data . . . . . . . . . . . . . . . . . . . . . . . . . . 21

8.5.2 dynamic data . . . . . . . . . . . . . . . . . . . . . . . . . 21

8.5.3 mockup data . . . . . . . . . . . . . . . . . . . . . . . . . 22

8.5.4 OWMap data . . . . . . . . . . . . . . . . . . . . . . . . . 22

8.6 Digital Ocean Backup Server . . . . . . . . . . . . . . . . . . . . 22

8.7 Deployment on EC2 Instance . . . . . . . . . . . . . . . . . . . . 25

9 Back-end development 25

9.1 Technologies . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

9.2 API Scrapers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25

9.2.1 Dublin Bikes Scraper : scrapdynamo.py . . . . . . . . . . 25

9.2.2 OWMap Scraper : open weather map scraper.py . . . . . 26

9.3 Flask Application . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

9.3.1 request static data . . . . . . . . . . . . . . . . . . . . . . 26

9.3.2 request info box . . . . . . . . . . . . . . . . . . . . . . . 26

9.3.3 request weather . . . . . . . . . . . . . . . . . . . . . . . . 26

9.3.4 get weather inuence . . . . . . . . . . . . . . . . . . . . . 27

9.3.5 request hourly prediction . . . . . . . . . . . . . . . . . . 27

9.3.6 request weekly prediction . . . . . . . . . . . . . . . . . . 27

9.4 Data Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

9.5 Accuracy Check . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

9.6 Error Logs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

10 Front-End development 30

10.1 Front-End technologies . . . . . . . . . . . . . . . . . . . . . . . . 31

10.2 HTML . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

10.3 CSS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

10.4 JavaScript . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32

10.5 JQuery . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32

3

11 Project Delivered 33

11.1 Key Features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

11.2 Performance optimisations . . . . . . . . . . . . . . . . . . . . . 34

11.3 Key Shortfalls and Planned Improvements . . . . . . . . . . . . . 34

12 LINKS: 34

4

1 Introduction

City biking has been traditional across Europe. Expanding cities like Copenhagen, Paris, Berlin, Vienna are often referred as bike friendly capitals. Bicycles are a healthy, environmental friendly way of commute. It needs simple infrastructure of a bicycle lane and bikes station network to operate. A motor road attempts to occupy garbage truck, bus, medium sized trucks, cars and motor bikes at a same time. Unlike this scenario, bicycle tracks carry single sized, single speed traffic providing congestion free commute.

Dublin bikes , which is operational since late 2006 has been a efficient commute option for Dublin citizens. It would be put on test more and more as Dublin is expanding. [City has seen population growth of 79,200 merely since 2015](https://www.google.com/publicdata/explore?ds=mo4pjipima872_&met_y=population&idim=subregion3:IE061&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met_y=population&scale_y=lin&ind_y=false&rdim=country_group&idim=subregion3:IE061&ifdim=country_group&tstart=1429052400000&tend=1555282800000&hl=en_US&dl=en&ind=false).

Aim of this project is development of an open working dynamic website which provides availability of bikes and bike stands at all Dublin Bike stations along with necessary statistics and weather related forecast to plan the journey.

This report aims to give the reader a rundown of the work and process that

went into the development of this application.

1. Aim

This Web Application is made to provide real time availability of bikes and bike stands at Dublin Bike stations along with weather condition at present time. Along with that, future availability at the station for following 5 hours. Given the Irish weather, Temperature and weather for corresponding time are also displayed. This application should be single page and easy to navigate. It should be useful for veteran and new users of Dublin Bike service to plan journey and schedule.

Figure 1: Team FFS-22

5

1. Project Management

This website is an instrument used to understand about Software development life cycle in a limited sized team using agile approach. This is 4 sprint project; logically divided into

planning and database setup

Front End development

Data analytics

Integration Testing, Deployment and Reporting

For project management, 3 tools were used :

Trello: Trello is Kanban board software by Atlassian. Trello provides flexibility to break up project into sprints and break it further into lists of cards. Facility of assigning label, story owner, checklists, attachments and deadlines to cards made role of scrum master efficient.

Github: Parallel development and CI/CD are key features of agile methodology. And version controlling avoids code conflicts, improve code coverage and reduces integration testing time.

This project had limited major threads, hence we preferred dedicated Trello list for communication instead of opting multichannel platforms like slack. Mobile phone application of Trello enabled conducting remote standups and sharing instant updates.

3 Sprint 1

Figure 2: Trello board for sprint 1

3.1 Group formation 12/02/19 9.00-11.00

3.2 Membership

Membership of the team was \_nalised, team members are Francesco Ensoli,

Fabio Magarelli and Se\_an Keyes. The group is number 22 and the supervi-

sor/customer is Karl Roe (teaching assistant).

3.3 Discussion

We discussed several concepts and examined the brief. We also examined some

existing websites and apps that perform a similar function.

6

3.4 Meeting 1; 14/02/19 9.00-11.00

3.4.1 Discussion

We discussed whether or not to write an SRS, examined existing applications

in more detail and discussed how to do the scraping as well as making some

mock diagrams of the user interface. Discussed what software we required RE

organization and coding.

3.4.2 Work done

\_ SCRUM environment set up

3.4.3 Work to-do

\_ Scraping software

\_ RDS

3.5 Meeting 2; 15/ 02/19 12.00-18.00

3.5.1 Discussion

This was our long Friday meeting. We discussed the workload and our approach

towards working as a group.

3.5.2 Work done

\_ Early version of the scraper

\_ Set up GitHub

\_ Set up virtual environments

\_ Created WhatsApp group for easier communication

3.5.3 Work to-do

\_ Fix scraper version 1

\_ Look into setting up RDS

3.6 Meting 3; 18/02/19 17.00-19.00

3.6.1 Discussion

We discussed the static and dynamic data and what needed to be in both.

We examined the new scraper version and discussed how it was to work. We

discussed the structure of the database and the tables in it.

7

3.6.2 Decisions made

We decided to add the banking variable to the static data. We also decided that

one table with data for all stops was superior to having the data divided into

one table per stop.

3.6.3 Work done

\_ Started the scraper

\_ Scraper started uploading data to the databases

3.7 Meeting 4; 19/02/19 9.00-11.00

3.7.1 Discussion

We discussed the user interface in greater depth, as well as discussing the use

of the weather information with regards to using this information to inform

the prediction. We also discussed the errors we were having when scraping the

JCDecaux API. We discussed what to do for the next meeting with Karl.

3.7.2 Decisions made

We decided we needed more sample designs for the user interface. We needed to

do the project backlog and to write the meetings up in a document for later use.

We also decided to create a second error log in order to store the information

about when stations don't update.

3.7.3 Work done

We examined our current documentation. We also looked at changes to the

GUI.

3.8 Meeting 4 20/02/19 13.00-13.30

3.8.1 Discussion

We discussed the work we did the previous evening. We looked at the Backlog

document and discussed the storage and functional uses of error data and the

format of the \_les in use. We discussed the packages used in the virtual en-

vironment online and the virtual environments we were each using personally.

We discussed the synchronization of the packages within. Documentation of

SRS-Like data.

We also discussed chartist.js for visualization of the graph data. We discussed

scrum platforms, vivify, jira, (you plan?)

3.8.2 To-do

\_ Finish current work; GUI, Documentation and Backlog

8

3.8.3 Wishlist additions

\_ Symchronisation of VIs with EC2 instance

\_ Station update analysis data

\_ Weather component of the analysis

\_ Ask Karl about the SRS-and other data.

3.9 Meeting 5 21/01/19 9.00-11.00

3.9.1 Discussion

we discussed the data and its potential uses eg. we could use the data to

help JCDecaux decide where they should install additional bikes and stands.

It would also help us see if the stands with banking access were more popular

then it could indicate that it would be good to add that functionality to other

stands. We discussed whether or not we should do a prediction over every 15

or 30 minutes instead of every 5, which would reduce the number of scrapes we

need, but make the the information less immediately useful.

3.9.2 Decisions made

Colours for icons, Scrape every 5 min to avoid inconsistency.

3.9.3 Work done

\_ Error reports in report not website

9

4 Sprint 2

Figure 3: Trello board for sprint 2

4.1 Meeting 6 26/02/19 9.00-11.00

4.1.1 Discussion

We discussed the main issues we faced in sprint 1 and the issues we would have

to tackle in sprint 2.

\_ The need to add a scrum master

\_ The burndown chart

\_ The backups for the database

\_ Error logs

\_ Youtrack

\_ Backlog

\_ Python code on ask to push into

10

\_ Weather data; do we scrape twice a day? Do we use historical weather

data for prediction or real time scraped data

\_ Wishlist idea; a section for a user to report a broken/faulty bike so that

the issue can be \_xed.

4.1.2 Work done

\_ Put all the things on Trello

4.2 Latest report

\_ Crash reports; AWS went down on the 28th of february

\_ error handling

\_ error messages from the programs if there is an issue is sent to all 3 mem-

bers of the team

\_ basic designs of the GUI

\_ Google maps API

\_ Dropdown menu search bar for individual stations

\_ Open Weather Map scraper

\_ backlog

\_ Error logs

\_ backup collected data

\_ look at historical vs actual open weather data.

4.3 Meeting 7/3/19

\_ We started trying to move the data on bike locations to the front end

using JQuery

\_ GUI work

\_ data analysis plan made, started working on \_guring out the code for it

11

5 Sprint 3

Figure 4: Trello board for sprint 3

5.1 Meeting 16/3/19

\_ |{We started trying to move the data on bike locations to the front end

using JQuery

\_ GUI work

\_ data analysis plan made, started working on \_guring out the code for it

12

6 Sprint 4

Figure 5: Trello board for sprint 4

6.1 Meeting 2/4/19

\_ discussed ask functionality and backlog items

6.2 Meeting 12/4/19

\_ We started trying to move the data on bike locations to the front end

using JQuery

\_ GUI work

\_ data analysis plan made, started working on \_guring out the code for it

13

7 Burndown Chart

The Burn-down Chart for the four sprint gives a rough estimate of the time spent

on the most important sections of the project. In respect to the progresses made,

the following surmise has been deducted. By looking at the chart, it is clear to

notice that the \_rst sprint in which the group got together, gave us the chance

to discuss on weaknesses and strengths of each team member, to organise and

decide which planned feature each person would have started to look into, hence

mostly planning took place. After the \_rst sprint, everyone adopted the agile

system con\_dently, and got busy working on their respective features, whilst

occasionally getting together during the allocated weekly practical times. By

the start of the third sprint, due to unforeseen circumstances, one of the team

members gradually disengaged himself, not committing to attending meeting

and provide updates on his part. This member was the one who's taken the

task of being the scrum master but also to work on the segment for the data

analytic for the predictions. It was by the start of the forth sprint that the other

members realised what was happening and started working on implementing

those features that were missing or not completed. It de\_nitely has taught

something important to the group, about the importance of simple but constant

communication and the essentiality of the scrum master \_gure.

14

Figure 6: Caption

15

8 Development Operations (Dev-ops)

Dublin Bikes web application project is end to end software development task involving database design and management, web scraping, data analytics and Machine learning, UI – UX design and cloud server deployment.

This section describe the Development operations decisions taken through the

entire project development.

8.1 Technologies

\_ (local) virtual environment

\_ GitHub repository

\_ Amazon AWS server

\_ Amazon RDS MySQL Database

\_ Deployment on EC2 Instance

8.2 Local machine

Project is developed using 3 distinct language categories :

Database : MySQL

Data scraping and API: Python, flask API

Front End : Java script, HTML, CSS

We used anaconda virtual environments for all 3 team PCs to eliminate problem with environment variables. Anaconda enabled us to quickly share updated environments avoiding package version problems. As development IDE, we settled for PyCharm, which is a good environment with large automation package library, and efficient GitHub integration.

8.3 GitHub Repository

We did set up 2 different Github repositories for this project, one for dynamic, and another for static development. Static development section consisted of web scraper and machine learning development efforts. Dynamic scenarios are maintained in another repo which involved website front end, deployed ML models and driver code. We appointed Engineer1 as master while Engineer2 and Engineer3 as collaborators. Master was tasked to asses quality and merge codes. Branching avoided problems associated with using single master branch and hard pushes.

Link: https://github.com/fabiom91/bikes dublin FFSTeam22

16

8.3.1 Repository Structure

8.4 Amazon AWS Server

As main cloud server we've used Amazon AWS EC2 free tier ubuntu instance.

While in the development phase has been set to accept connections only via

SSH, this has been changed to accept all connections on deployment. The

EC2 Server runs in multiplexing (using tmux) the Dublin Bikes API scraper,

the Open Weather Map API scraper, the Accuracy evaluation script and the

Flask server application. In the last months of development we got 2 server

crashes which were most likely due to AWS: the \_rst one occurred on the 28th

of February at 16.00 but speaking with other teams, this crash occurred to

everyone else. The second crash has not been tracked correctly. After the \_rst

17

crash we agreed to implement a function that would have sent an email to each

one of the team members when a crash occurs. The actual function works well to

catch crashes due to API Scrapers malfunctions and Database connection errors

but is not e\_ective to catch server crashes. Maybe a further implementation on

the backup server would have been able to catch them but this feature has not

been implemented due to time restrictions.

8.4.1 Custom Packages installed

Python 3 and Conda shall be installed on the EC2 Instance.

\_ pip list

{ ask 1.0.2

{ matplotlib 3.0.2

{ mysql-connector-python 8.0.15

{ numpy 1.15.4

{ pandas 0.24.0

{ pip 18.1

{ pypyodbc 1.3.4

{ requests 2.21.0

\_ brew list

{ wget

{ mysql

{ tmux

18

19

8.5 Amazon RDS MySQL Database

The RDS Database have been developed mostly using MySQL Workbench. All

the relations of the single database have been created manually by the devel-

opers and populated dynamically through python scripts and functions. The

initial Database schema was composed by three tables: one for the static data

from the Dublin Bikes API, one for the dynamic ones and the last one for test-

ing. While the testing table has been left into the database till the end of the

deployment, one other table have been added to store info regarding the Open

Weather Map API scraper. We discussed regarding the opportunity to save data

with a certain frequency and which data were meaningful for our application.

This is the \_nal Database Schema:

20

8.5.1 static data

Contain static informations regarding all the stations in dublin city. Since those

informations does not change in the short time, they can be updated as needed

(even just once).

number (PK), name, address, lat, lng, banking

8.5.2 dynamic data

Contain dynamic informations regarding all the stations in dublin city. Those

data are updated every 5 minutes by the scraper running on the EC2 Instance.

Since number and last update are primary keys, the relation does not accept

duplicates tuples in which both station number and time already exist. Fur-

thermore, number is a foreign key of the static data relation "number" such as

the relation accept data only if they matches a known station.

number (PK) (FK), status, available bike stands,

available bikes, last update (PK)

21

8.5.3 mockup data

Contain Dynamic data for practice. It's been used when implementing new

features or components in order to preserve the actual data from unexpected

changes. This table is been populated with the dynamic data of 24h with a

scraping interval of 10 minutes.

number (PK), status, available bike stands,

available bikes, last update (PK).

8.5.4 OWMap data

Contain a timestamp as primary key and a weather condition such as rainy,

sunny etc. It is populated with the main value of the weather scraped from the

Open Weather Map API with an interval of 1h.

timestamp (PK), weather (NN).

8.6 Digital Ocean Backup Server

Database backups are made on csv \_les. The database is checked for new en-

tries every hour on a separate server and every 1000 new records regarding

dynamic data table are saved into a backup \_le while a single \_le contain the

OWMap data. Due to space limitation, a dedicated server has been created on

Digital Ocean in order to store the backups. At this stage there is no other

strategy regarding the deletion of old data. All data are stored physically on

the backup server as csv \_les.

Di\_erent solutions have been attempt in order to make regular and exep-

tional database backups during the development: At \_rst we tried exporting

the database with MySQL Workbench but we encounter a compatibility issue

with mysql version and the mysqldump one which we weren't able to solve in a

practical way so we discard this option.

22

After this \_rst attempt, we implement a python function running on a local

machine which query the database and store the data received into 3 csv \_les:

one for each table. We had 2 problems with this solutions: the query of the

dynamic data table took more than an entire night to compute and the csv

\_le was in the order of the 10Gb which wasn't acceptable for any of the local

machine of the team members.

The third solution, the one currently working, is not the best solution nor a

de\_nitive one but for now it suits the team needs: A new server instance has

been created on Digital Ocean and on this instance, a python script has been

uploaded and run in a mutliplexer which query the dynamic data table of the

database for the last 1000 entries and save them in a separate csv \_le. The

Server solution takes advantage of the computation power of the cloud server

and its big hard drive to store the data until a de\_nitive solution will be found.

(such as backing up only the last n number of rows etc.)

Since the size of the backup is quite prohibitive, there is no way to actually

show it. If you need a demonstration, ask a Fabio Magarelli to show you the

backup server running (before the end of May, then the server will be de\_nitely

shut down). This is a screenshot of the backup server:

23

24

8.7 Deployment on EC2 Instance

Important: Due to a technical issue with pip on the EC2 server, we had to

temporarily deactivate all cache features from the python code. The issue can-

not be resolved in time prior the submission of the product. This issue is due

to a pip package problem since the functools.lru cache packet is now in-

corporated into the latest versions of python3 and shouldn't require any install

through pip.

Deployment has been made on the EC2 instance modifying the default ip and

port: 127.0.0.1:5000 to 0.0.0.0:80 and setting the security group of the

Server to allow any connection. Other solutions such as nginx and Gunicorn

have been considered but discarded due to operational reason. Furthermore, we

are not much concerned about eventual security breaches at this stage. the link

to the running web application is :

http://ec2-34-238-40-161.compute-1.amazonaws.com

9 Back-end development

This section describe the Back End Development decisions taken through the

entire project development.

9.1 Technologies

\_ API Scrapers

\_ Flask Application

\_ Data Analysis

\_ Accuracy Check

\_ Error Logs

9.2 API Scrapers

API Scrapers are essentially python programs to make some API calls to retreive

data from the Dublin Bikes API and the Open Weather Map

9.2.1 Dublin Bikes Scraper : scrapdynamo.py

JCDecaux API python scraper algorithm, get the data from JCDecaux and save

the relevant ones (dynamic) in the RDS Database. Loop every 5 minutes. It

also handle errors such as:

\_ station update missing : write the station number and the timestamp

into logs/station error.csv .

25

\_ station update missing : write the station number and the timestamp

into logs/station error.csv .

\_ API call errors : write the error log into logs/error logs.txt and send an

email to the team!.

\_ DB Connection errors : write the error log into logs/error logs.txt and

send an email to the team!.

9.2.2 OWMap Scraper : open weather map scraper.py

Open Weather API scraper, get the data re weather conditions in Dublin City:

gets only the "main" realtime forecast: rainy, sunny etc. and save it in the

RDS Database. Loop every 1 hour. It also handle errors re API calls errors and

database connection errors. save those errors into logs/error logs.txt. and send

an email to the team!

9.3 Flask Application

Our web application use Flask to compute all the main computations and con-

nect to the various APIs and the database. In particular, the the following are

the functions implemented in the ask app (main.py):

9.3.1 request static data

request the static data from the database to populate the google map with the

stations.

9.3.2 request info box

request the last update of number of bikes and stands for a given station num-

ber in order to display them in an info box on a station marker click on the map.

Get as argument the station number and return the available bikes/stands and

banking at the moment (The last available update)

argument(s) = station number

return = station number : f available bikes : value , available stands : value ,

banking : value g

9.3.3 request weather

Request real time weather from OWMap API direct call.

Return the real time (now) weather from OpenWeather Map API to display

over the map For the user to know what is the weather now.

return : weather dict = f'temp' : value, 'description' : value, 'wind speed' :

value, 'icon link' : valueg

26

9.3.4 get weather inuence

compare for every station the mean of available bikes and available bike stands

when the weather is "clear" and when the weather is the current weather (get

this info from OWMap API). return the percentage of di\_erence between the 2

predictions.

return : inuence = fbikes: value, stands: valueg { e.g. value : 12% more/less

9.3.5 request hourly prediction

This prediction is made for a selected station returning the average number of

available bikes and stands for a given day, grouped by hour.

Get as argument the number of the station and the day of the week (Monday,

tuesday. . . ) and return a JSON \_le with the prediction of the average hourly

availability of bikes/stands along the day

argument(s) : station number, week day

return : h pred = ftime : [values], bikes : [values], stands : [values]g

9.3.6 request weekly prediction

This prediction return the average availability of bikes and stands for a selected

station grouping them by day of the week.

Get as argument the station number of which we want to know the prediction

and return the prediction of the average daily availability of bikes/stands along

the week

argument(s) : station number

return : w pred = fday : [values], bikes : [values], stands : [values]g

9.4 Data Analysis

When implemented the Flask functions: request weekly prediction,

request hourly prediction and get weather influence we used the aver-

age availability of bikes and stands for given input(s) to make a \_rst data anal-

ysis to predict the future availability of bikes and stands. This method is not

the most accurate one. We considered to use a linear regression model but due

to time constraints this option has been discarded.

9.5 Accuracy Check

Since we couldn't implement a linear regression model in time for submission,

an accuracy study has been performed on our "model" which use the average

availability of bikes and stands in the past to predict the future one. The results

can be found in this repository in accuracy log/accuracy.csv. The accuracy

log currently show an accuracy of the 20%. The accuracy of the prediction

27

is calculated for the hourly prediction comparing the predicted values for each

hour against the actual ones.

9.6 Error Logs

As already mentioned, errors regarding API and MySQL connection and station

update into dynamic data are catched by the various functions in through the

application back-end. All those error logs are saved in di\_erent \_les:

\_ error logs.txt : contain all the eventual logs of API Connection and DB

Connection errors occurring from the scrapdynamo.py script run. Those

error logs are very important because they can cause the scraper to crash

which may result in a loss of data.

\_ station error.csv : this csv \_le contain informations regarding errors

that occur when a station data is not updated: since the number of station

and the last update are foreign keys of the relation, the database cannot

accept tuples in which both of those values are duplicates of a previous

tuple. This kind of errors does not make the scraper crash but the record

is skipped. It could be relevant to collect data regarding the un-collected

records in order to monitor the functionality of the JCDecaux service.

\_ ask error.txt : contain all the eventual logs of errors due to a failure

connecting the web app to the data analysis python scripts through ask.

This is important to detect and correct eventual bugs. Note: since a

proper model for data analysis is not been implemented, this \_le is not

currently catching any errors.

Furthermore, errors logs due to API and MySQL connection are sent via email

to all the members of the team. Note: the sender email address for the error

logs (daftscraping@gmail.com) is an email already used for other applications

since this would be seen only by the team members there was no need to create

a new email address for this purpose.

28

29

Figure 7: Architecture structure

10 Front-End development

This section explains the technologies used for the front-end development along

with the structure of the \_les deployed. Having used Flask throughout the

testing phase, the structure required by Flask to work is as follow:

\_ bikes dublin FFSTeam22

\_ index.html

{ Static

\_ CSS

\_ images

The main HTML \_le is inside the main folder, meanwhile the css \_les and

the images used are respectively inside the CSS and images folders.

30

10.1 Front-End technologies

\_ HTML

\_ CSS

\_ JavaScipt

\_ JQuery

10.2 HTML

For the project we have used the HTML5 standard, which has made easier to

structure an entire page, relegating most of the styling to CSS. Utilizing newer

elements like "header" to create the uppermost page DIV and "footer" for the

lowermost one. But also canvas used on the charts DIVs and several other

elements like calendar and Geolocation, which can track the user position in

real time if required, making the whole process very easy to implement. The

whole page is divided into four main sections, namely the header, the map, the

information window and the footer.

\_ Header

\_ Map

\_ Information window

\_ Footer

The header is where we have decided to show the weather predictions coming

from the weather API. Below it there is the map which takes up most of the

screen estate. The map we have used is from the Google Map API. The map can

be customized to ones preference and has very well written manuals to further

customize it and implement on all of its functionalities. Following the map there

is the section where all the information is displayed. To welcome the user and

avoid cluttering the screen with much information at the start, there is a splash

screen covering the section. The idea is for the user to only see the general

sensible info and later select the extra information required. Upon selecting one

of the markers, the splash screen disappears and the actual window replaces

it. For the purpose of this application the footer is left empty. In a future

update information like the person to contact or the product owner name could

be added if wanted.

10.3 CSS

CSS3 was used to stylize in the web page the arrangement of the elements.

There are two \_les for the CSS:

/.static/css/grid.css

/.static/css/style.css

Throughout the page elements have been displayed using `Grid' layout.

While inside sections the `Flex box module layout is also used. Everything in

31

relation to modularity and display arrangement has been created and modi\_ed

within the `grid.css' \_le. For instance the header width spacing has been done

using ex grid, which makes the header responsive and keeps the ratio between

all the elements included. The second \_le 'style.css' is the one where all the

style options have been done: from the size, to the color of di\_erent elements,

including info-windows. The style that was chosen for the web-page vaguely

resembles that of an arcade video game, in term of colors and proportions.

10.4 JavaScript

JavaScript is the main engine behind the site, transforming the originally static

page into a dynamic one, increasing the client interactions. It is used to make

requests to the scraper, and stylize the various DIVs in which the information is

contained or sent to. For the project the Javascript code resides in the HTML

\_le. We have used some vanilla code to implement some functions, and some

Jquery code to do the pull requests to the server. Javascript o\_ers many frame-

works. We have used Jquery extensively (described next) and also ApexCharts

to visualize the charts on the site. These charts are already implemented to

animate and be dynamic, adapting to the div they are placed in. Furthermore

using a JSON style format is it possible to customise them to anyone's liking.

10.5 JQuery

This JS framework is used mostly for the Ajax capabilities of easily sending

pull requests to the server and for the way it simplify complex methods into few

lines of code. For the site we used two versions one for the standard framework

and one for the jQuery-UI theme implementation.

The jQuery has several functions relative to the Flask application, that pull

speci\_c data to \_ll di\_erent sections of the website.

Speci\_cally:

request static data: this function returns all the information about the stands,

like name, number, status, location and if it has banking available. These infor-

mation is then used to create all the markers on the map, with corresponding

variables.

Furthermore by knowing the location of each station using its latitude and lon-

gitude, it is also possible to calculate the distance with the user and advice

on the nearest stations. Within the request is also possible to create a div to

show the information pulled from the server to the user, in this case the nearest

stations from the user.

request info box: is used to request to the scraper the speci\_c information about

a stand. It returns the latest update on the number of bikes and stands available

and displays it in an info-window that is created within the function

32

request hourly prediction and request weekly prediction: functions are both

sending the required data to the respective chart, placed at the bottom of the

screen. The charts are respectively showing the amount of bikes and stands

available throughout the day and on a weekly prediction.

get weather inuence: function returns the predictions calculated based on

the weather for the day selected. These predictions are returned on the initial

splash screen.

In the next example we can see how a request is made and variable is passed.

The example shows how the request is made to retrieve the station data, to fur-

ther manipulate it.

11 Project Delivered

The web app interface provide sensible information about the current status

of the Dublin Bike service, with advanced functionalities to make predictions

based on weather forecast.

11.1 Key Features

\_ Display real time availability of bikes and stands.

\_ Display real time weather forecast.

\_ Predict hourly and weekly availability for bikes and stands.

\_ Display current position of the user (hard coded).

\_ Display nearest stations based on the user location.

\_ Display weather inuence on bike and stands availability.

33

11.2 Performance optimisations

To improve performance when retrieving data from the database, we imple-

mented on the Python built-in LRU-cache library. Due to the deployment oper-

ational issues (described in the dev-ops section) this feature has been temporary

deactivated.

Another issue encountered when deploying the system on the server, has

been with the HTML Geo-location feature, which requires a SSL certi\_cate to

operate. Being unable to retrieve a certi\_cate before the deadline, we have been

advised to hard-code the user location just to show proof the functionality.

11.3 Key Shortfalls and Planned Improvements

Due to unforeseen circumstances with one of the team members, some of the

features that were originally planned, have been delay or postponed to a fu-

ture update of the app. These features are: Data analysis models (e.g Linear

regression), proper deployment methods (e.g usng NGINX), Geolocation and

SSL certi\_cates, security improvments of the server and the database, proper

analisys of station that systematically fail to update using station error.csv

\_le, proper accuracy analysis of the predictions based on accuracy.csv \_le.

Other considerations that have been made during the development phase, were:

to provide the user with a registration \_eld, which would grant access to a user

area, where is possible to report issues like faulty stations, faulty bikes with a

direct line of contact in case of any other issue related to the service.

12 LINKS:

GITHUB repository for the project

https://github.com/fabiom91/bikes dublin FFSTeam22

Project Web page

http://ec2-34-238-40-161.compute-1.amazonaws.com/