Data Science projects

Hamza Imloul

January 3, 2022

1 Introduction

These projects are developed using R or/and Python. To view tutorials, user guides, and further documentation, please visit my portfolio.

2 Projects

2.1 Decision support tool for the renewal of equipment.

Saved operational and maintenance costs.

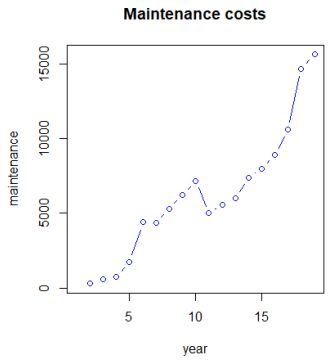

Figure 1: Reported waste in the transportation process.

2.2 Data pre-processing tool, of geofencing events dataset using R

Data preparation, is the most important part of the data analysis process, increasing value from data. Importing data set, understand the business problem, and select variables in the output.

Using Rest API calls to request data from servers of both platforms, it is the best approach since the data received is raw, in a JSON format, and not consuming storage resources. But requires a specific knowledge of HTML request and Restful APIs structure.

Helpfully, a useful method to manage and keep the work organized is to work with geofencing alerts since the two platforms send them by e-mail. And by that way, the mailbox can be considered a centralized fleet management platform specifically for the pickup, travel and delivery operations.
The inconvenient of this technic is that every geofencing alert mail is incorporating a descriptive text and image, what consume storage and wireless network data. In addition, data retrieving demands an appropriate program to connect with mailbox provider server, and the use of regular expressions to extract the needed information from every mail. But for now, it still the most affordable technic, since it provides us the needed and real-time updated information.

The number exceeds the speed limit.
Rate of call responsiveness.
The number of times when the driver started after the planned starting time in a group.
The number of days off in a month.
Exceeding driving time limit (9 hours).
Brutal acceleration.
Brutal braking.

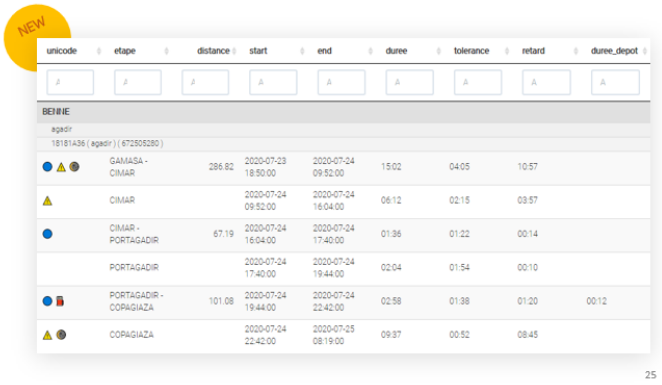

Figure 2: Reported waste in the transportation process.

2.3 Predicted the behavior of the transportation process

First you have to upload the image file from your computer using the upload link in the file-tree menu.

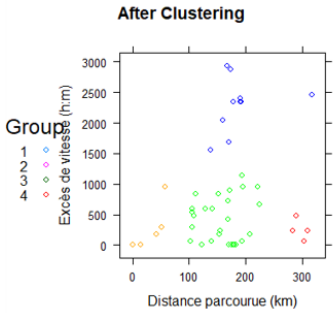

Figure 3: Reported waste in the transportation process.

2.4 A GPS Based solution, for the dynamic assignment problem

Logistics touches every part of a business and even goes beyond to include suppliers, carriers and customers. Needless to say, the logistics are complicated with many interrelated components. This high degree of complexity is difficult to manage without the use of models that faithfully represent the processes and their interactions. Reduce costs and downtime. In this Article, we will apply Hungarian Algorithm, which is an optimization algorithm.

Modelling the problem After appliying design research, some If you are not familiar with trasnportation problems, you can check this good website.

Algorithm We use the Hungarian algorithm, Assign n missions to n available trucks E_{rv} set of available real vehicles.

E_{dv} = set of dummy vehicles.
E_s = set of sources of missions.
E_d = set of destinations of missions.

Cost matrix

$$c_{vpij} = cf_v(d_{pi} + d_{ij}) + ca_{ij} - p_{ij} \tag{1}$$

where

c_{vpij} = the cost to assign the truck v located in p to the task i to j.
cf_v = cost per km of fuel consumption for the truck v.
ca_d = cost per km of the resource (driver) allocation.
d_{ij} = distance traveled to perform task from the source i to the destination j.
d_{pi} = distance traveled to perform task from the truck position p to the source i.
b_{ij} = bonus of driver to performs the mission i to j.
p_{ij} = profit of transportation from the source i to the destination j.

the $vpij^{th}$ element is the cost of assigning the resource v located in p, to the task: source i to destination j. the matrix is 4-dimensionnal. We research to add penalty cost related to in-site time in sources and destinations, to avoid missions with a large cycle time. + Probability to blockage InSite due to infraction (wheels state...).

2.5 Modelled the parcels' routing time.

Problem: calculate the delay J+i of delivery of a package in order to obtain the number of engaged packages to be delivered within J+i.

Inputs: datetime-depot, abb-depot, abb-destination, id-package.
data sets : intra-trips, inter-trips, Delays-matrix, Available-Trucks-capacity, packages-details.

2.6 Designed a R-Shiny App to monitor the transportation activity

You can make lists with automatic numbering ...
The first movement detected in a day of all the vehicles visualized in a spatial map. Visualization technics: coloring the circles based on the time of the first movement.
Tracking the movement of all vehicles.
Evolution of distance traveled along a period of time.
Evolution of driving time.

2.7 Machine Learning Analysis with Venue Review Data in Calgary

For this second task, we would like you to analyse a dataset that contains review data of different venues in the city of Calgary, Canada. With the help of several machine learning techniques that we have learnt in the course, you will be tasked to distill insights from this social media dataset.

Two of its notable features are the geocoding of every reviewed venues and the availability of a considerable amount of text data in it, which lend to its ability to be processed using spatial and text analysis techniques respectively.
As a prelude to the analysis prompts below, have a brief think about some of these questions:

What can we discover about the venue review data?
Are there any spatial patterns that can be extracted from the data?
Can we build a machine learning model that predicts review rating for unseen data points using the text of the reviews?

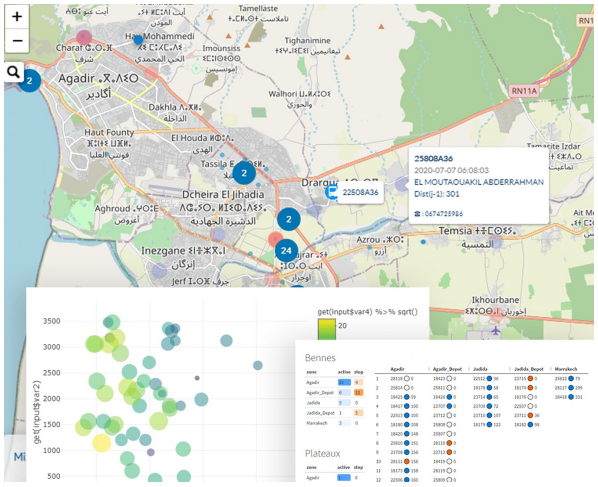

Figure 4: Tracking of vehicles status on map.

2.8 Sentiment Analysis using NLTK

Run a lexicon-based sentiment analysis (eg. NLTK Vader Sentiment Analyser) on the textual data, then report and discuss the results. Does the lexicon sentiment score associate with the venue ratings provided by the users?

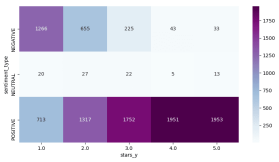

Figure 5: Reported waste in the transportation process.

2.9 Capacitated Pickup & Delivery Vehicle Routing of taxis service

Usually the template you're using will have the page margins and paper size set correctly for that use-case. For example, if you're using a journal article template provided by the journal publisher, that template will be formatted according to their requirements. In these cases, it's best not to alter the margins directly.

Figure 6: Reported waste in the transportation process.

2.10 Anomaly detection tool exposing the wasting time in the process

ddddd

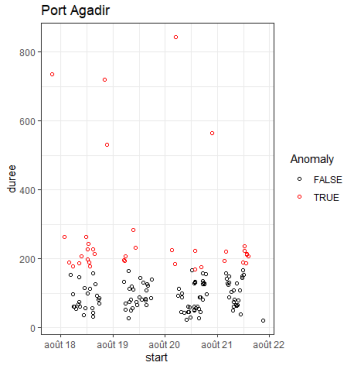

Figure 7: Reported waste in the transportation process.

2.11 Stop ξ Search data analysis to predict crime outcome in London

If you have an upgraded account.

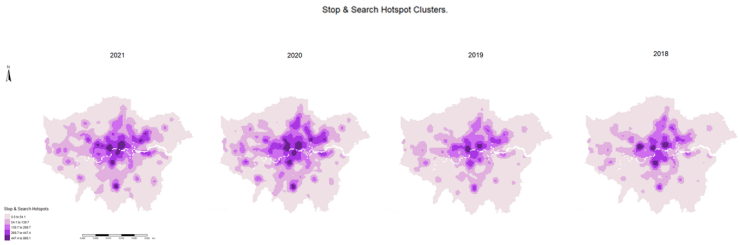

Figure 8: Reported waste in the transportation process.

2.12 Data Analysis of the NanoString assay

Bro that's good

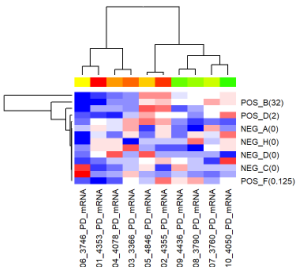

Figure 9: Reported waste in the transportation process.

2.13 Closeness centrality

Hamd

Figure 10: Reported waste in the transportation process.

route from A to B

Figure 11: Reported waste in the transportation process.

2.14 Optimized the routing of sales persons through schools in california

Real problem case.

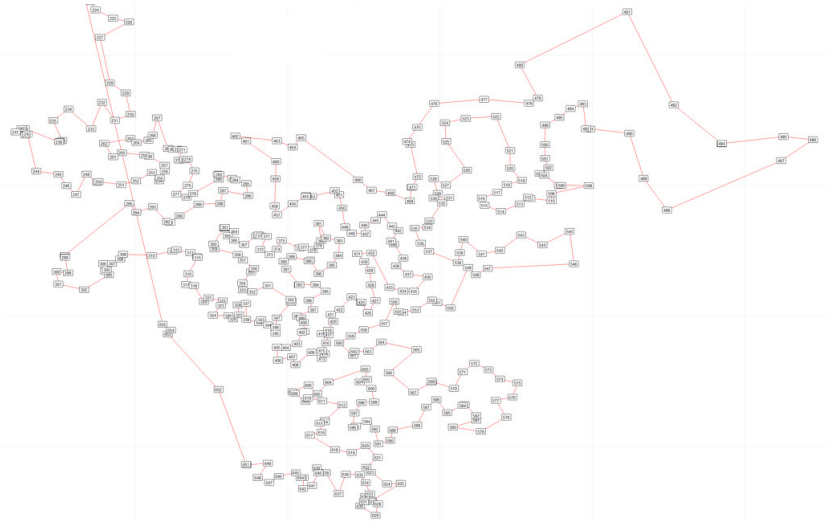

Figure 12: Extracted the itinerary through 9800 points.

2.15 Designed an app visualize the locations of a research center members

Real problem case.

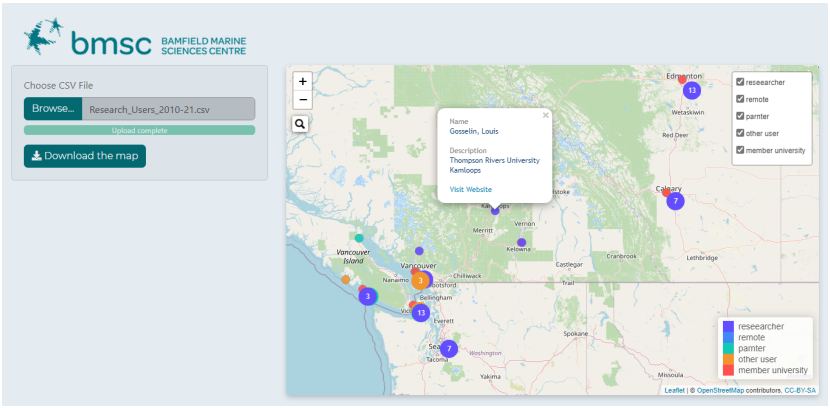

Figure 13: R Shiny app.

2.16 Designed an interactive app visualize the weather data

Real problem case.

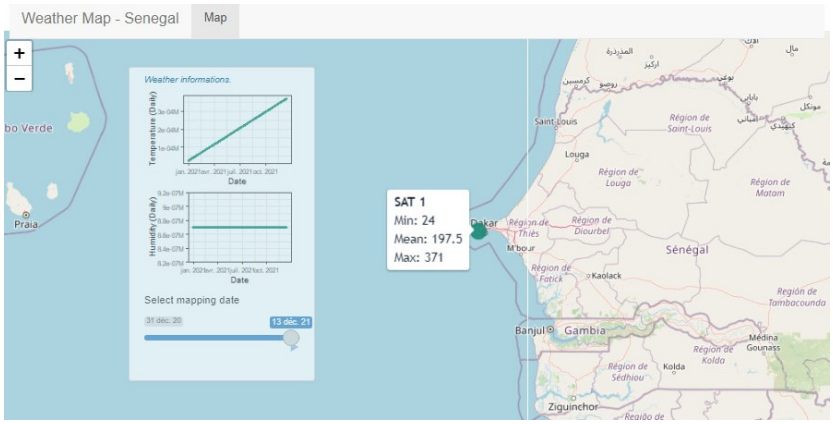

Figure 14: R Shiny app.

2.17 Filling automation of delivery paper sheets

Real problem case.

Entering exit vouchers in a database is important since it allows the transport operations carried out to be justified. In this document, we propose a new input method, automating this task, and analyzing the advantages of this method.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	date	n_bon	societe	site	transporteur	vehicule	conducteur	cin	code	quantite	libéllé	poids_vide	heure_date_e	poids_brut	heure_date_s
2	10/07/2020	3004052705	Ciments du Maroc	USINE AIT BAKA	MIXTRA	25807A36	SAID EL ABID	QA111352	130053	25700	Coke de pétrole brut MTS	17		43	10072020 114701
3	09/07/2020	3004052569	Ciments du Maroc	USINE AIT BAHA	MIXTRA	25807A36	SAID EL ABID	QA111352	130053	26020	Coke de pétrole brut MTS	17		43	09072020 223406
4	09/07/2020	3004052358	Ciments du Maroc	USINE AIT BAHA	MIXTRA	25811A36	ABDELHADI SISSIN	JT50761	130053	26060	Coke de pétrole brut MTS	17		43	09072020 153028
5	10/07/2020	3004052668	Ciments du Maroc	USINE AIT BAHA	MIXTRA	25811A36	ABDELHADI SISSIN	JT50761	130053	26340	Coke de pétrole brut MTS	17		44	10072020 101956

Figure 15: R Shiny app.

2.18 Designed a Power BI app for retail data

Real problem case.

Entering exit vouchers in a database is important since it allows the transport operations carried out to be justified. In this document, we propose a new input method, automating this task, and analyzing the advantages of this method.

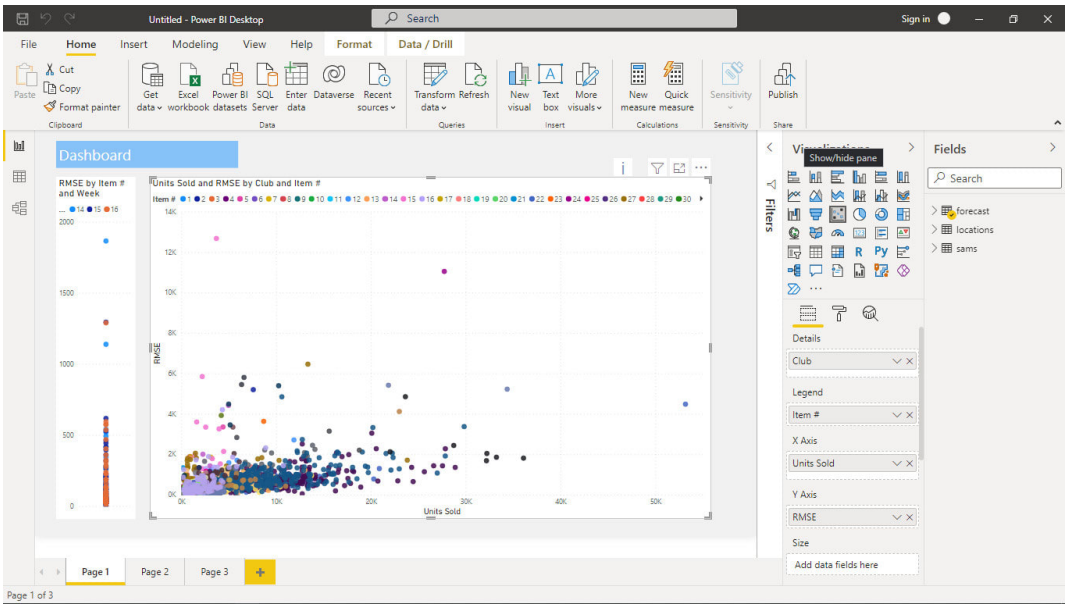

Figure 16: Power BI app.

We hope you find Overleaf useful.

References