

Mahmoud TRIGUI « Data Scientist »

Street El Ain Km 6, Environment Boulevard
3042 Sfax - TUNISIA

Mobile phone TN : (+216) 97 304 501

E-mail trigui.mahmoud7@gmail.com

Self-tape <https://youtu.be/OUIjBuMxoXU>

LinkedIn www.linkedin.com/in/mahmoudtrigui

Skill IQ <https://app.pluralsight.com/profile/mahmoud-trigui>

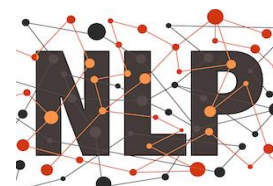
Acclaim www.youracclaim.com/users/mahmoud-trigui/badges

Zindi https://zindi.africa/users/Mahmoud_Trigui/competitions

Personality <https://www.16personalities.com/profiles/46e44008a0245>



analytic.



Titled with a statistics and data analysis engineering degree with a solid problem solving thanks to my mathematics-physics background, having +9 years of experience, passionate about anything related to machine learning, loving dealing with structured data, exploring data-mining insights for our daily issues tabular data, finding and creating new features through detailed analysis and approaching projects with different perspective

Key Words

Analysis, ANOVA Analysis, Business, Clustering Relationship, Construction, Churn, CVM, CRISP-DM, Data storytelling, Dataiku, Decision rules, Dimensionality Reduction, Embedding, Feature engineering, Forecasting, Geospace Analysis, GPT-4, Hypothesis, IBM Watson, Image conversion, LLM, Methodologies, MLForecast, Network analysis, NLP, Oracle DB, Presentation, Profiling, Pydantic, R Markdown, SAS-GUIDE, SAS E-MINER, Sampling Techniques, Shiny App, Segmentation, Spatial Data, Transformations, Time Series Decomposition, Tutoring, VBA

ZiND!



[See Mahmoud_Trigui's top challenges](#)

Contents

Data Scientist @ Sofrecom Tunisia (part of Orange Group)	3
PDF Info Extraction with LLM	3
Time Series Forecasting for Business Demand	3
CoFinancing Prediction for FTTH Deployment	4
Data Scientist @ Kiota Intelligence	4
Shiny App	5
RMarkdown Report	5
Exploratory Data Analysis	7
Matching Script	8
Clustering Relationship	8
Data Scientist/Analyst @ Tunisie Telecom	9
Client Segmentation	10
Multi-Sim Detection	13
Data-Mining Process	14
Family Communities	15
Ad-hoc Analysis / Reports	15
Data Science Projects @ Freelance	18
Facebook Page Analysis	18
Vehicle Weight Estimation	18
Data Science Tutor	19
Research Statistician @ LAAS-CNRS	22
Alarm Automatic management	22
Machine Learning Competitor @ Zindi	24
Spatial Data	24
Various	26
Assessments, Courses & Skills	28
Pluralsight	28
LinkedIn	28
Acclaim	29
DataCamp	29
Coursera	30
edX	30
Stanford Online	31
MIT ProX, Udacity & DataQuest	32
+ Personality	34

Data Scientist @ Sofrecom Tunisia (part of Orange Group)

PDF Info Extraction with LLM

We developed a PDF Audit Information Extraction System using **GPT-4** that automatically processes multi-page audit reports. The system converts PDFs to images rather than text because some information are already on logos and graphics, it also search a specific items in a pre-defined reference dictionary.

Built with a modular Python architecture, it features custom **Pydantic** validation models, optimized **image conversion**, and Base64 encoding for **LLM** compatibility.

The solution generates structured JSON outputs while maintaining data integrity through comprehensive validation mechanisms, significantly reducing manual processing time for audit professionals.

```
...
Processing PDF: /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit.pdf
Successfully converted /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit.pdf to 12 images
Results saved to: /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit_V1.json

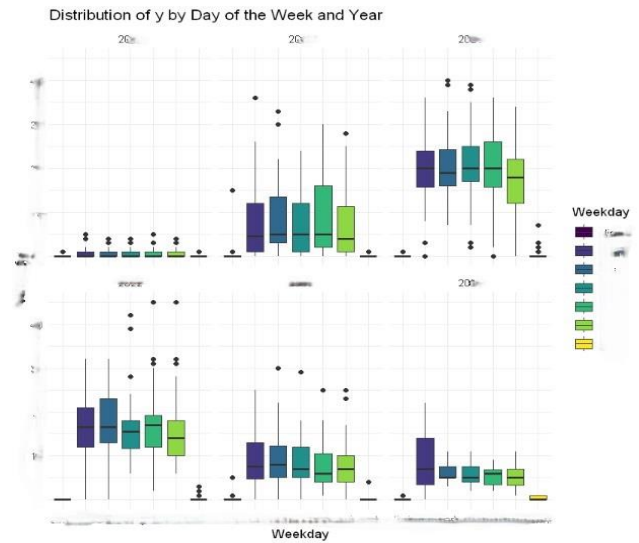
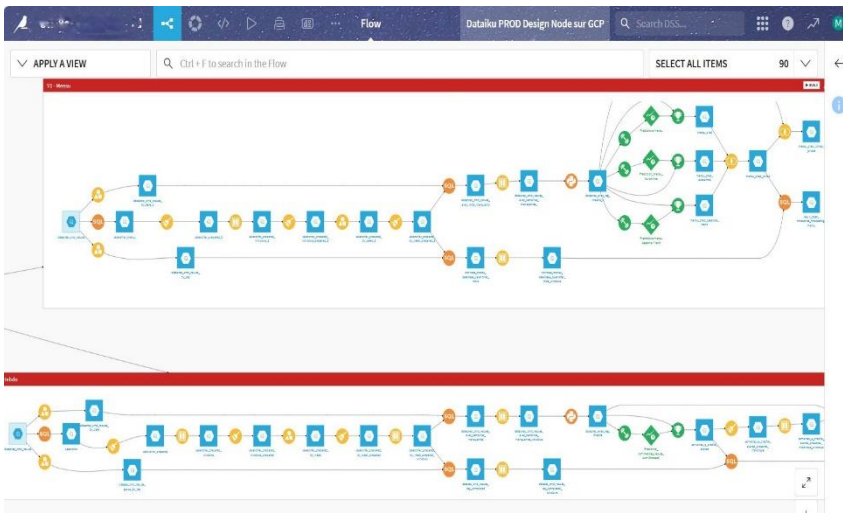
Processing PDF: /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit.pdf
Successfully converted /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit.pdf to 14 images
Results saved to: /home/jupyter/python-code/resources/audit/Audit_Report_ID_100600051_Audit_Communit_V1.json
```

Time Series Forecasting for Business Demand

The goal is to generate 12-month & 53-week demand predictions at multiple business levels (department & offer). The system addresses the challenge of limited historical data through advanced feature engineering, incorporating holiday calendars, weekend patterns, and seasonal decomposition techniques.

My implementation leverages R's specialized **forecasting** packages (feasts, fable) to create an ensemble of models including ARIMA, ETS, Holt-Winters, TBATS, Prophet and Deep learning. The solution achieves approximately 5% of error.

Key technical components include time series **decomposition** (STL, AutoSTR) to separate trend and seasonality patterns.

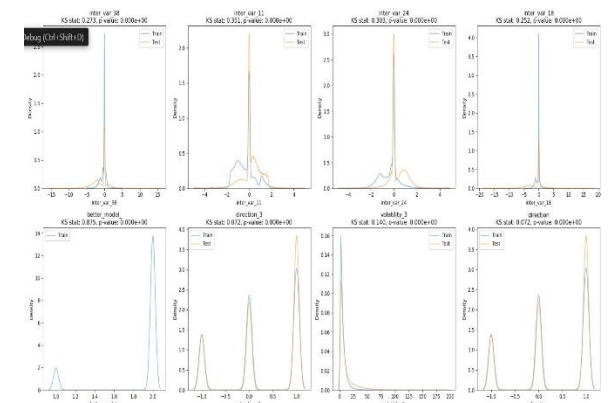
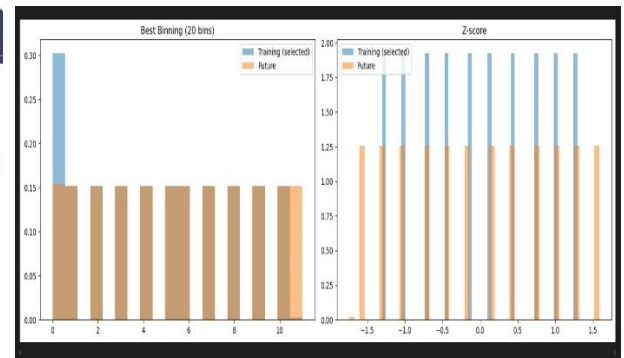
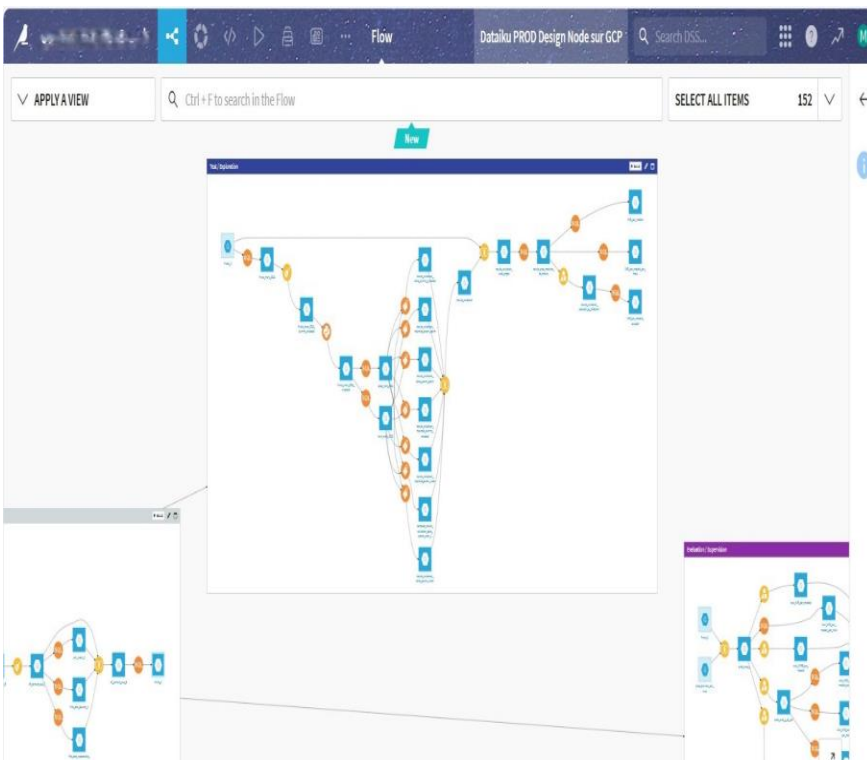


CoFinancing Prediction for FTTH Deployment

The goal is to predict co-financing rates for FTTH. The project involved migrating Python scripts to production-grade **Dataiku** workflows, prioritizing SQL and native components for performance optimization, and designing an efficient Analytics Base Table creation process.

For model enhancement, we validated **MLForecast** function effectiveness, tested additional predictive variables, and evaluated alternative modeling approaches including other time series models and transforming the problem into a regression task.

The part of feature engineering did improved the model, including creating interactions terms and doing the necessary **transformations** to have similar distributions across different period time.

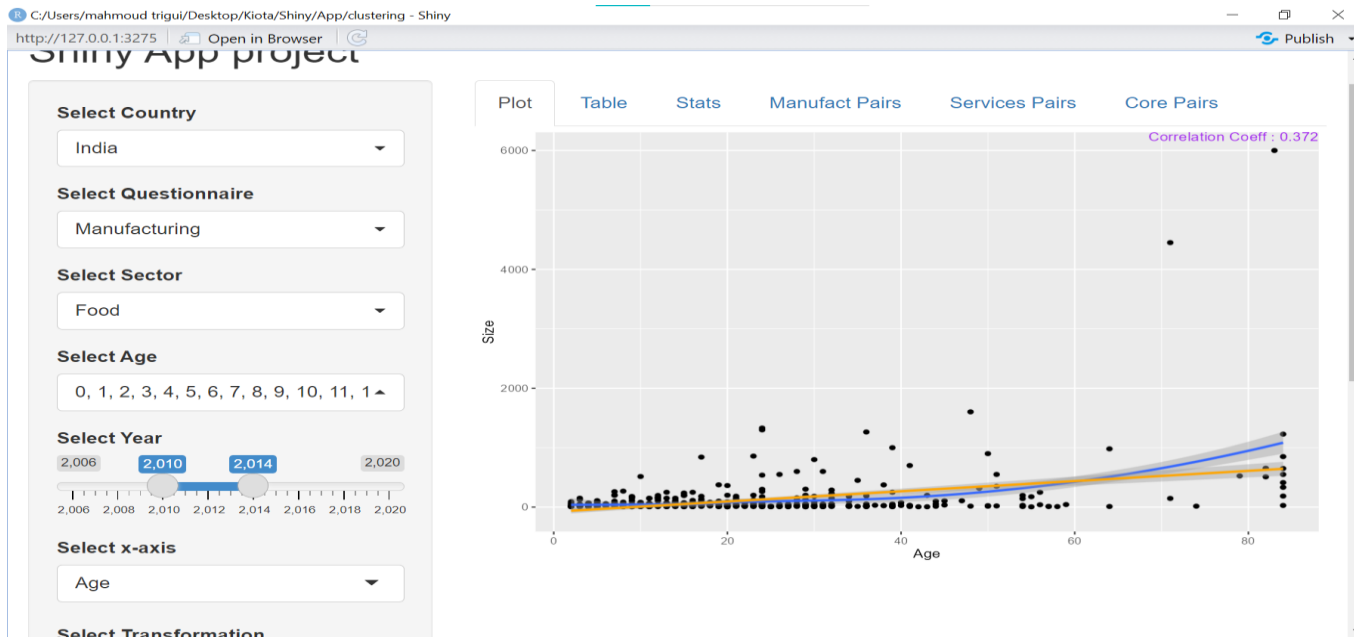


Data Scientist @ Kiota Intelligence

My main work was to create the desired output for our endpoints with plumber, all type of task in data science with R included. All data were investment and funding rounds related.

Shiny App

The goal was to provide a tool that allow clients to have statistics and graphs about their data-entry based on many filters.



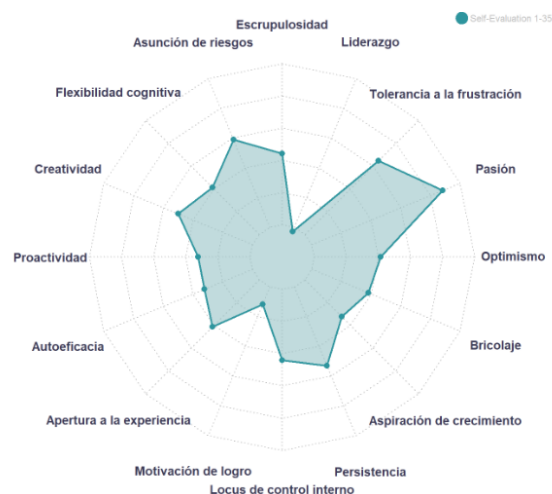
R Markdown Report

Generate a complete psychometric report about entrepreneurs and teams

Executive Summary

According to our analysis of your questionnaire, We found that you have **average scores in several of the factors evaluated**. This is Consider **it convenient for your entrepreneurship**, because:

- You generate ideas through knowledge integration belonging to different social and cultural domains, facilitating the development of flexible reasoning and the creation of creative and constructive ideas.
- You experience positive feelings throughout the entrepreneurial process and you make your purposes make you feel identified and motivated, allowing you to keep constant your interest and effort to achieve what you you propose.
- You are willing to create ideas until you find the most suitable without rush to take one to avoid the discomfort of experiencing ambiguous situations, favoring more decision-making planned and reasonable.





KIOTA BRAIN ANALYSIS

Evaluation certificate

Copyright Kiota 2022 all rights reserved



Intelligence Artificial deductive



¡Nueva herramienta!

Persistente

0.4 %

Estratégico

0.23 %

Organizado

0.16 %

Competente

0.14 %

Enérgico

0.08 %

WHAT DO WE DO?

We analyze five qualities that entrepreneurs tend to have more successful through this novel model of artificial intelligence deductive.

Based on the lexical analysis of your description, we calculate the level that you present in each of the qualities.

INTERPRETATION

According to the results, we can estimate that the qualities that the more they describe to you are those of being a **Persistent and Strategic**.

It is worth mentioning that these qualities are closely related to traits. entrepreneurs such as (factors that will be detailed later).



Psychometric details: PLANNING

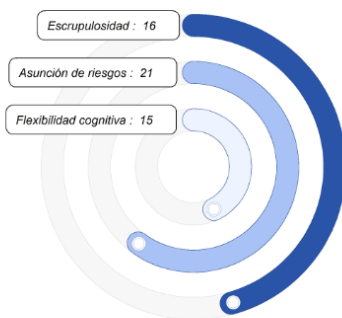
Evaluation certificate

Copyright Kiota 2022 all rights reserved



Scrupulousness

16 | MEDIO



WHAT DOES IT MEASURE?

Conscientiousness is the ability of a individual for being organized, responsible and meticulous with all Matters related to documentation and establishment of objectives of the entrepreneurship. Conscientiousness has a positive relationship with the locus control, openness to experience and these, in turn, were associated positively with entrepreneurial intent and success. Are positive effects on the creation of a business by the Responsibility and self-control of the entrepreneur.

INTERPRETATION: You have obtained an average score in scrupulousness. These results indicate that you are characterized by levels means of responsibility, self-control and entrepreneurial discipline. So regular, you are detail-oriented and organized with the related matters of your entrepreneurship, reflected in the fact that you assume commitments and propose objectives of your business in the medium or long term, although sometimes you It makes it difficult to plan ahead for your tasks.



Assumption of risks

WHAT DOES IT MEASURE? Risk taking is capacity inherent in an individual making risky decisions and taking on risks in uncertain or

INTERPRETATION: You have obtained an average score in risk-taking. These results suggest that you are a person. that maintains an analytical profile in the

EDA

The goal is to have a descriptive statistics and a test hypotheses between different stages in proportion.

```
{
  "stages": {
    "applications received": [
      {
        "stage": "applications received",
        "id": "62aaed798e05ed3a8066889e",
        "title": {
          "en": "In which vertical of the program your startup fits better?",
          "es": "¿En qué vertical crees que encaja tu startup?"
        },
        "type": "categorical",
        "result": [
          {
            "factor": "",
            "amount": 2,
            "percentage": 40
          },
          {
            "factor": "Nuevas proteínas",
            "amount": 1,
            "percentage": 20
          }
        ]
      }
    ]
  }
}
```

```
    {
      "amount": 1,
      "percentage": 50
    },
    {
      "factor": "Sostenibilidad",
      "amount": 1,
      "percentage": 50
    }
  ]
},
"interpretation": "There is a significant difference in proportion between the
two datasets over the applications received stage and the interview stage",
"test": "Chi-Square",
"significant_variable": [
  {
    "variable": "Nuevas proteínas",
    "deviation": "Has changed from 20% in applications received to 0% in
interview. \n"
  },
  {
    "variable": "Salud y bienestar",

```

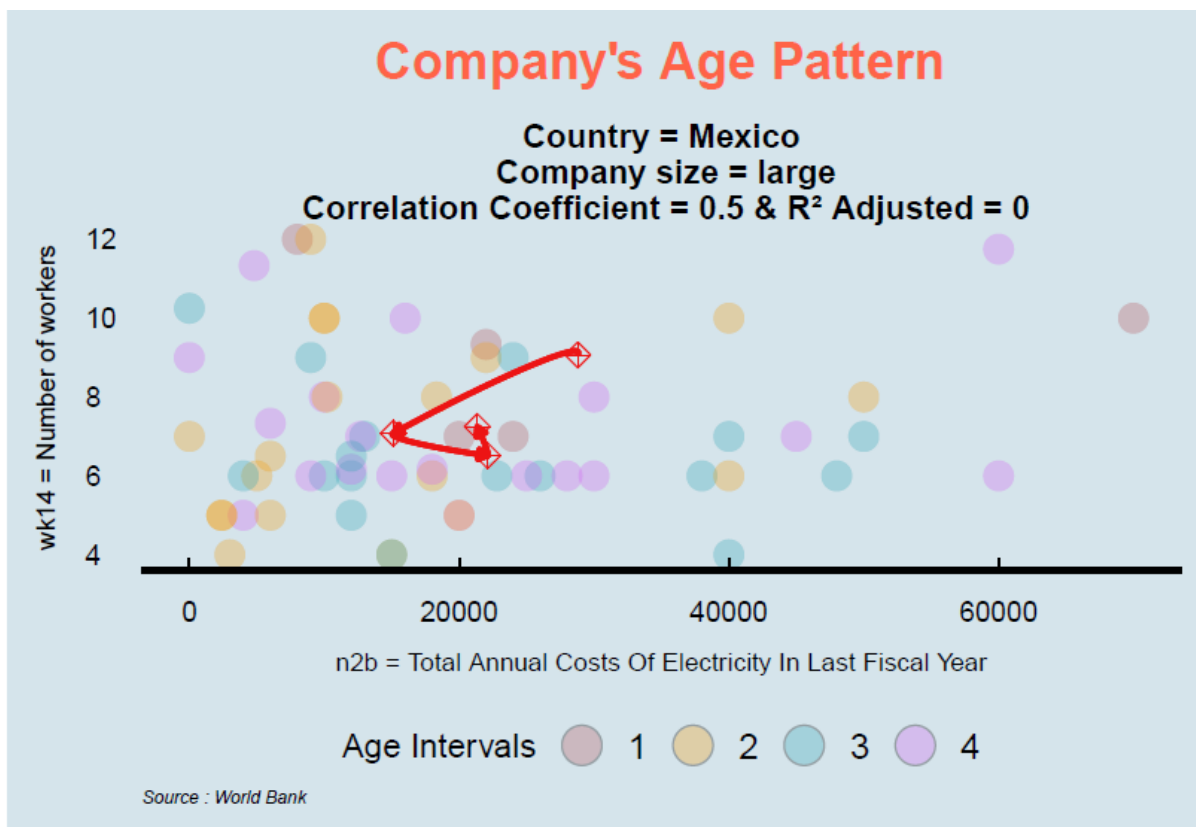
Matching Script

The goal of this project was to connect the right investors to the right entrepreneurs based on many filters.

```
[  
  {  
    "investors_name": "100x.VC",  
    "investors_id": "6",  
    "investors_nlg": "We propose this investor because according to his investment  
      thesis, he invests in companies at early stage,pre-seed stage"  
  },  
  {  
    "investors_name": "25madison",  
    "investors_id": "27",  
    "investors_nlg": "We propose this investor because according to his investment  
      thesis, he invests in companies at early stage,pre-seed stage"  
  },  
  {  
    "investors_name": "500 ventures"  
  }  
]
```

Clustering Relationship

The goal of this project was to find any potential relation between different variables and to connect their centroids to see the evolution over time.



Data Scientist/Analyst @ Tunisia Telecom Group

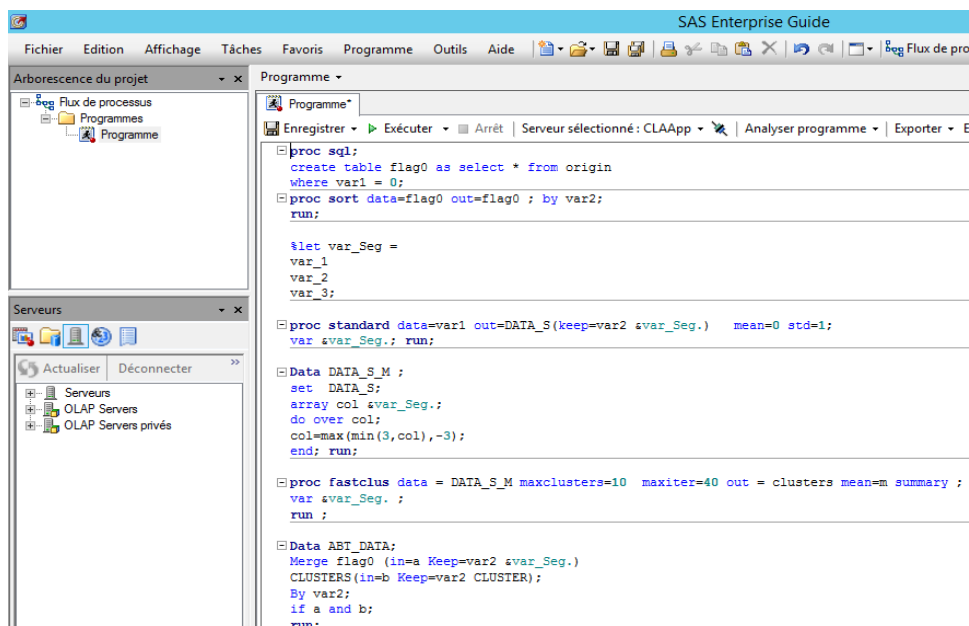
I had the opportunity to work with big amount of data, millions of rows including hundreds of variables describing clients through different source and tables.

A large amount of time for the data-mining project was dedicated to the ABT **construction** via:

- ✓ Data reliability-check
- ✓ Data gathering
- ✓ Data cleaning

So, in order to accomplish these tasks, I need to have a provident coding in SQL for Oracle (Toad), in addition to the basic programming of SAS and its functionalities in SAS Guide.

```
SELECT DISTINCT
  NUMERO_APPEL, var, 0.25 * var3 + 0.25 * var2 AS score
FROM (SELECT DISTINCT
      NUMERO_APPEL,
      CASE
        WHEN (var5 = 1 OR Forwarding = 1 OR var6 = 1)
        THEN 1 ELSE 0 END AS var3,
      FROM table_1 s
      WHERE imei IN
        ( SELECT DISTINCT IMEI
          FROM table_1 s
          INNER JOIN
            (SELECT DISTINCT var
             FROM table_4
             WHERE var7 IN ('A', 'S')) d
            ON d.var = SUBSTR (s.id, 4, 8)
          GROUP BY IMEI
          HAVING COUNT (IMEI) > 1 AND IMEI IS NOT NULL) d0
      ON s.NUMERO_APPEL = SUBSTR (d0.id, 4, 8)
    LEFT JOIN
      (SELECT DISTINCT APPELANT, var8
       FROM table_2
       WHERE ( ( mois IN ('01', '02') AND annee = '2018') OR (mois = '12' AND annee = '2017'))
       ON s.NUMERO_APPEL = 10.APPELANT
    LEFT JOIN
      (SELECT DISTINCT id, var_4
       FROM table_3) l
      ON s.NUMERO_APPEL = l.id
      (s.MOIS IN ('01', '02') AND s.ANNEE = '2018')
      AND SUBSTR (s.NUMERO_APPEL, 1, 2) <> '91'
    GROUP BY s.NUMERO_APPEL);
```



Client Segmentation



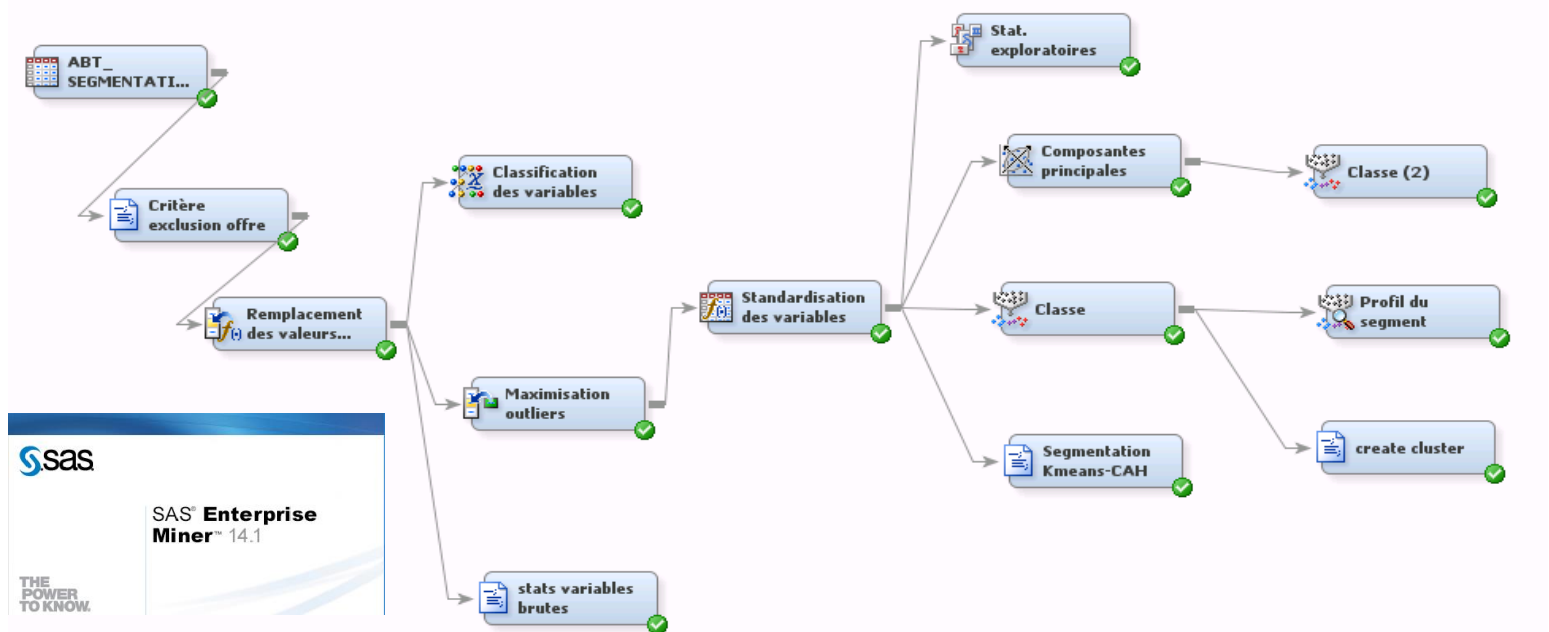
The business needed a new segmentation for our biggest part, the prepaid client base, to use it in their targeting campaigns, in the **CVM** and in various reports requests.

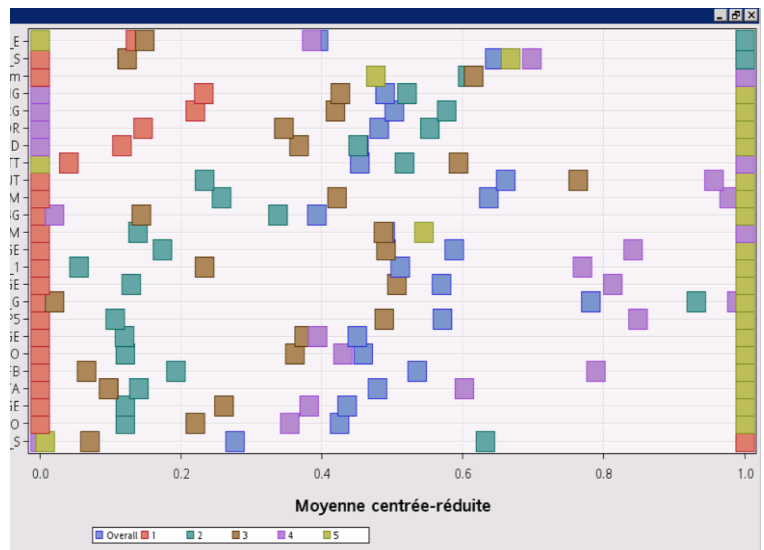
The initial phase was the data preparation along with the schedule and the specification & mapping document.

Taches du projet	Minimum	Moyenne	Maximum
1 - Compréhension de la problématique et fouille de données	1	2	3
2 - Cahier de spec	1	2	3
3 - Mapping	1	2	3
4 - Création de la table cible et les variables dimension	1	2	3
5 - Création des tables usages trafic	3	4	5
6 - Traitement et analyse des variables trafic	1	2	3
7 - Création des tables facturation	3	4	5
8 - Traitement et analyse des variables facturation	1	2	3
9 - Création des variables pour mieux expliqué le modèle	1	2	3
10 - Création de l'ABT final	1	2	3
11 - Etude prédictif de l'ABT	2	3	4
12 - Mise en œuvre des algorithmes de scoring OU segmentation	1	2	3
13 - Élaboration du modèle ou segmentation	3	4	5
14 - Amélioration du modèle ou segmentation	1	2	3
15 - Validation du modèle	1	2	3
16 - Traitement et analyse du résultat	2	3	4
17 - Présentation des résultats	2	3	4
Période global	26	43	60

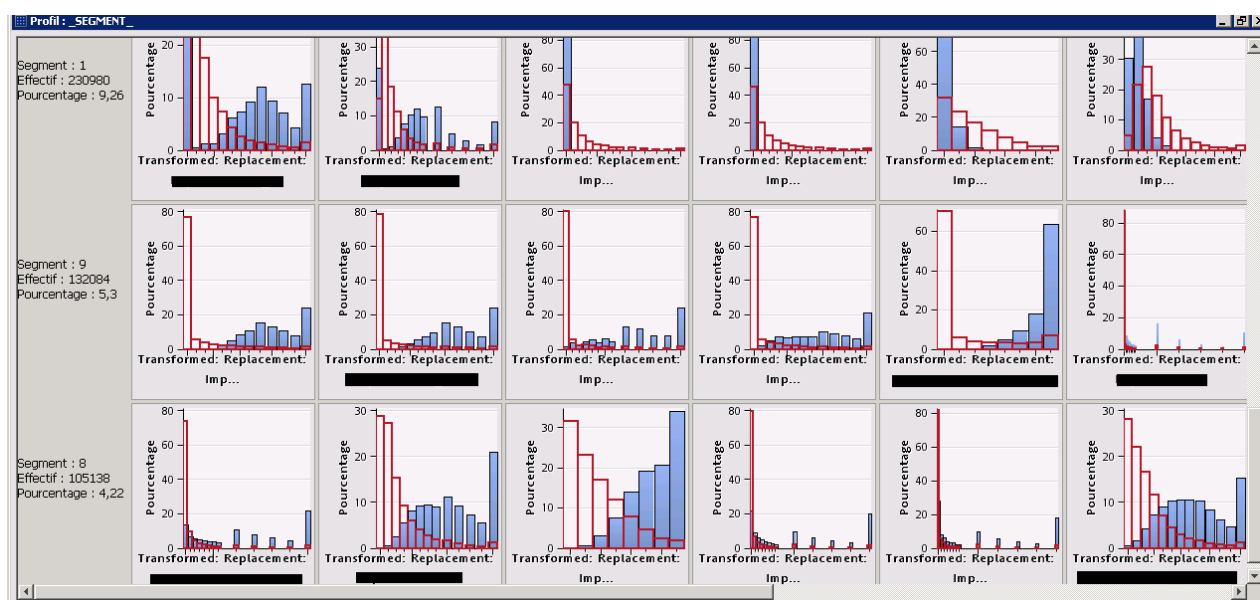
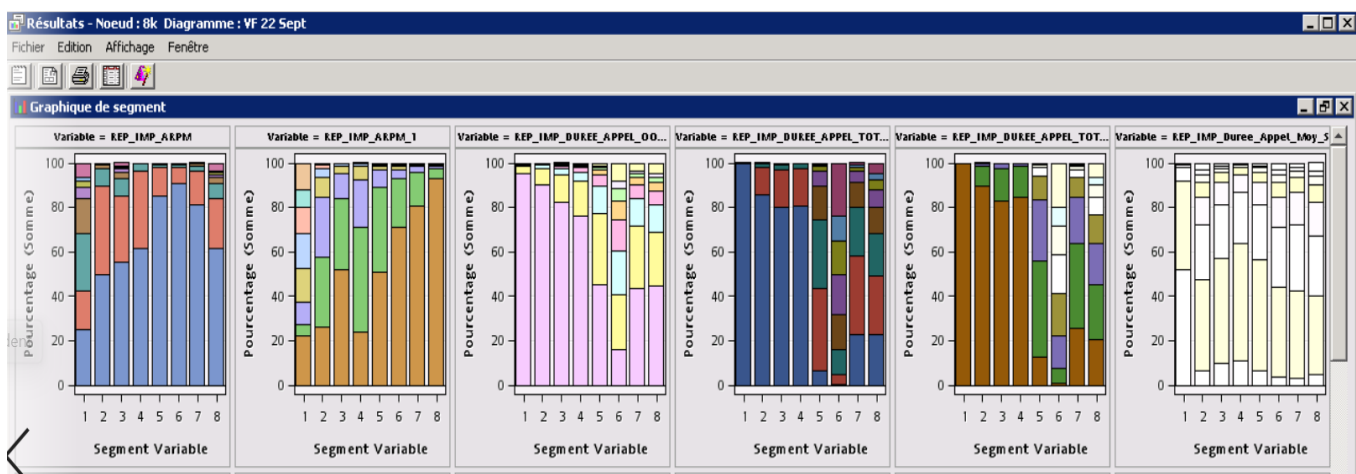
1 INTRODUCTION	3
1.1 CADRE 3	
1.2 OBJECTIFS	4
1.3 IDEE 4	
2 SPECIFICATION DES CRITERES ET MAPPING	5
2.1 FICHE MODELE	5
2.2 MAPPING DE L'ABT	5
3 TRAITEMENT DES DONNES	37
3.1 NCEUD FILTRER	37
3.2 NCEUD IMPUTER	52
3.3 NCEUD REMPLACER	52
4 CLUSTERING	57
4.1 LES VARIABLES SELECTIONNEES	57
4.2 SAS E-MINER	59
4.2 RESULTATS DU CLUSTERING	59
5 PROFILING	57
5.1 PROFILING DETAILLE EXCEL	57
5.2 PROFILING SAS	59
6 SEGMENTATION	57
6.1 REGLES ET SEGMENTATION	57

Phase 2 was the important one: trying different **segmentation** methods, different **methodologies** (one client per row or one client per period). Work done with **SAS E-MINER**.





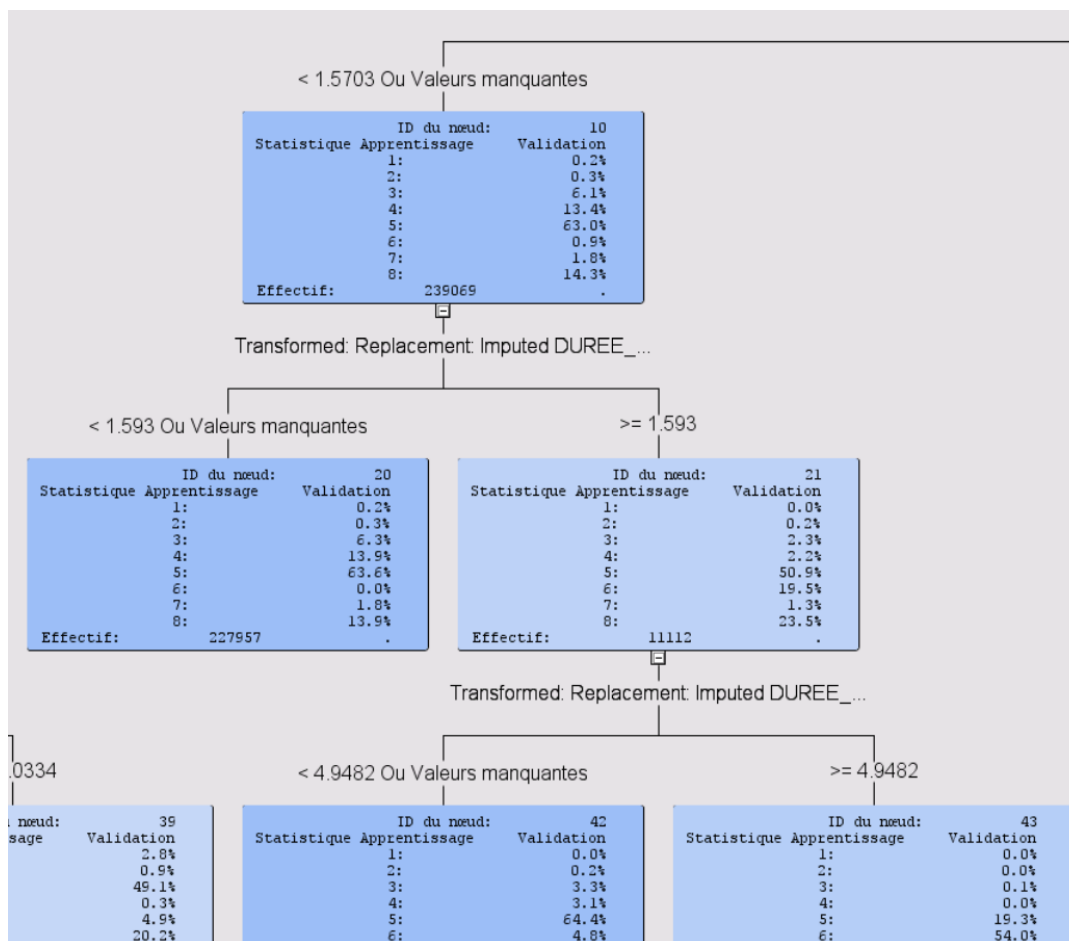
Primordial step to validate our clustering result is the detailed **profiling**, which will be affected by the business strategies and goals of the department.



1	Variable	New	SEG 8	SEG 7	SEG 6	SEG 5	SEG 4	SEG 3	SEG 2	SEG 1
236	HANDESET									
237	***	1.83	0.55	1.00	-0.08	0.04	-0.28	-0.13	-0.01	-0.21
238	***	0.65	0.07	0.11	0.01	0.01	-0.04	-0.01	0.02	-0.06
239	***	1.45	0.14	0.04	0.06	-0.07	0.05	0.02	0.00	-0.10
240	TRANSFERT									
241	***	0.41	0.44	0.37	0.13	-0.03	-0.11	0.01	-0.22	-0.01
242	***	0.25	0.17	-0.13	0.04	0.09	-0.13	0.23	-0.17	-0.07
243	***	0.22	0.25	0.24	0.00	-0.09	-0.04	0.00	-0.11	0.03
244	***	0.13	0.11	-0.05	0.05	0.04	-0.05	0.07	-0.08	0.00
245	RATIO									
246	***	0.08	-0.02	-0.06	-0.02	-0.02	-0.01	0.00	-0.01	0.08
247	***	1.04	-0.02	-0.06	0.03	0.12	-0.10	0.02	0.02	-0.02
248	***	2.59	-0.50	-0.09	4.79	-0.21	-0.88	-0.40	1.46	-1.27
249	***	0.46	-0.26	-0.33	0.37	-0.14	-0.27	-0.22	0.17	0.62
250	***	3.40	4.62	0.24	-0.46	-1.57	-0.26	-2.57	5.80	-1.27
251	***	4.10	6.14	1.39	-0.28	-1.71	-0.16	-2.33	4.66	-1.54
252	EVOLUTION									
253	***	9.82	10.61	9.95	18.11	2.66	0.97	-5.16	-5.02	-5.88
254	***	13.70	14.66	12.84	24.45	3.95	1.40	-6.54	-6.74	-8.67
255	***	4.34	1.41	1.58	4.75	4.01	0.02	-1.85	-1.66	-2.17
256	***	-4.72	-3.80	-2.76	-6.07	-1.39	0.15	1.47	1.39	2.24
257	***	4.97	3.97	2.75	6.57	1.46	-0.19	-1.57	-1.40	-2.31
258	***	6.09	4.67	-2.32	17.07	3.84	2.21	-3.17	-3.39	-3.72

Phase 3 was for the supervised task: once the clusters are finally reviewed and validated, we build a model to assign every customer of each new month to its segment.

Here we took a different approach in order to make it easier for the non-technical audience and for the directors, not by applying simply a classification model but to try to set easy **decision rules** extracted from the previous decision tree-based one, and then with the help of Tableau or Excel, we fit the final groups.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Cluster	Priorité	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var_9	Var_10	Var_11	Var_12
2	C1 : Segment_1	2	>= 0.335	335	>= 15.9208	>= 81.6007	>= 21.0371	>= 23.9323						
3	C2 : Segment_2	5	< 0.01	< 0.03			>= 9.6748		>= 11.9688		< 3.515		>= 177.3908	
4	C3 : Segment_3	3		>= 0.395			< 24.7854							
5	C4 : Segment_4	1		< 0.005		< 3407.182	>= 12.0078			>= 8.2737	>= 3.5792			
6	C5 : Segment_5	4	>= 0.735				< 17.194						< 567.72	
7	C6 : Segment_6	8	< 0.015	< 0.225			< 13.5088					>= 1.3229		>= 1.3229
8	C7 : Segment_7	6	< 0.035	< 0.365								< 1.0278		>= 46.165
9	C8 : Segment_8	7	< 0.015		< 0.4458		< 12.8315					< 1.0093		< 46.165

Multi-Sim Detection



We had to do several analytics to build initial **hypothesis** for our model since we do not have any record of multi-sim users in our databases, the idea was to proceed with clustering but with different approach (the goal to identify even small groups or who had a very practical behavior).

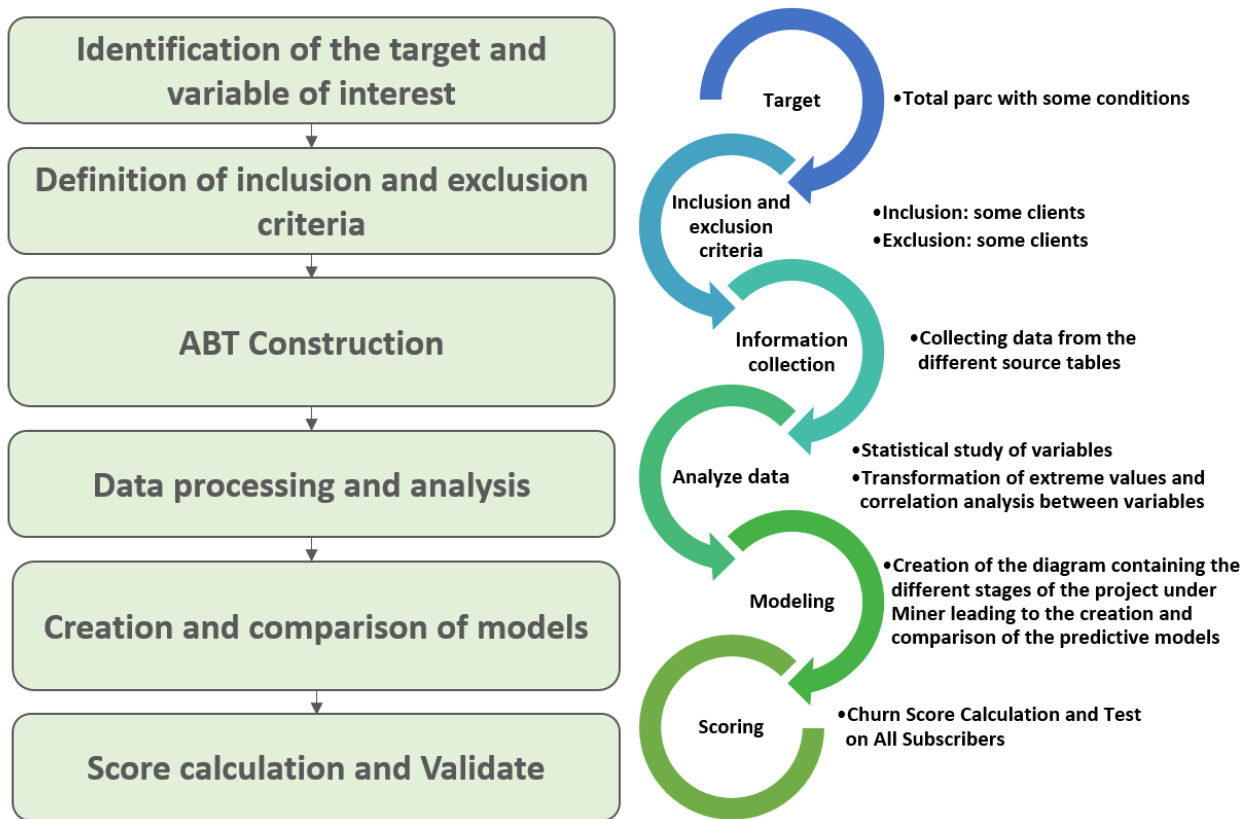
With a detailed profiling, we were able to distinguish the most suspicious group, validate this by doing a call-center campaign, then build a predictive model on top of that data (with **sampling** techniques).

	A	B	C	D	E	F
1		MultiSim_HV	Clus_2	HV	Clus_4	DATA
59	MULTISIM					
60	***	479,800	720,214	546,963	196,589	100,769
61	***	646,500	738,530	548,402	196,664	104,417
62	***	464,866	712,981	546,315	196,766	100,570
63	***	573,208	728,264	546,686	196,049	103,639
64	***	400,080	692,878	544,037	195,821	97,268
65	RECHARGE					
66	***	371,439	703,895	539,820	196,932	92,318
67	***	37,701	113,512	93,676	73,248	18,913
68	***	306,946	666,377	530,161	196,827	88,841
69	***	120,752	421,629	380,104	189,793	61,727
70	***	152,907	228,025	179,415	96,799	46,258
71	***	635,357	713,414	534,300	193,201	97,595
72	INTER					
73	***	484,709	713,738	543,082	195,244	99,846
74	***	67,640	121,306	134,077	50,462	25,918
75	***	66,525	87,761	88,242	27,983	16,860
76	SMS					
77	***	237,244	424,342	403,891	143,491	76,372
78	***	128,059	273,584	309,337	113,504	59,833
79	***	140,823	290,831	316,767	109,642	61,268
80	***	136,325	283,367	312,565	107,938	60,448
81	***	20,216	49,332	60,990	21,950	12,360

CRISP-DM

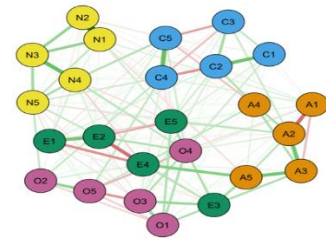


I was integrated partially in the **churn** project among other ones, but every **data-mining process** is like:



		3%	9%	6%	13.1%	16.5%	21%	15%	16%
		Segment_1	Segment_2	Segment_3	Segment_4	Segment_5	Segment_6	Segment_7	Segment_8
Old churn score	VLR	28%	31%	60%	66%	51%	29%	22%	28%
	LR	59%	60%	35%	23%	41%	28%	26%	17%
	MR	2%	2%	1%	1%	1%	5%	6%	4%
	HR	7%	5%	2%	6%	4%	20%	27%	19%
	VHR	4%	2%	2%	4%	2%	17%	20%	31%
New churn score	VLR	91%	92%	96%	92%	94%	65%	53%	51%
	LR	3%	3%	2%	4%	3%	12%	15%	15%
	MR	2%	2%	1%	2%	2%	10%	14%	14%
	HR	1%	1%	1%	1%	1%	7%	10%	11%
	VHR	2%	1%	1%	1%	1%	5%	8%	8%

Family Communities



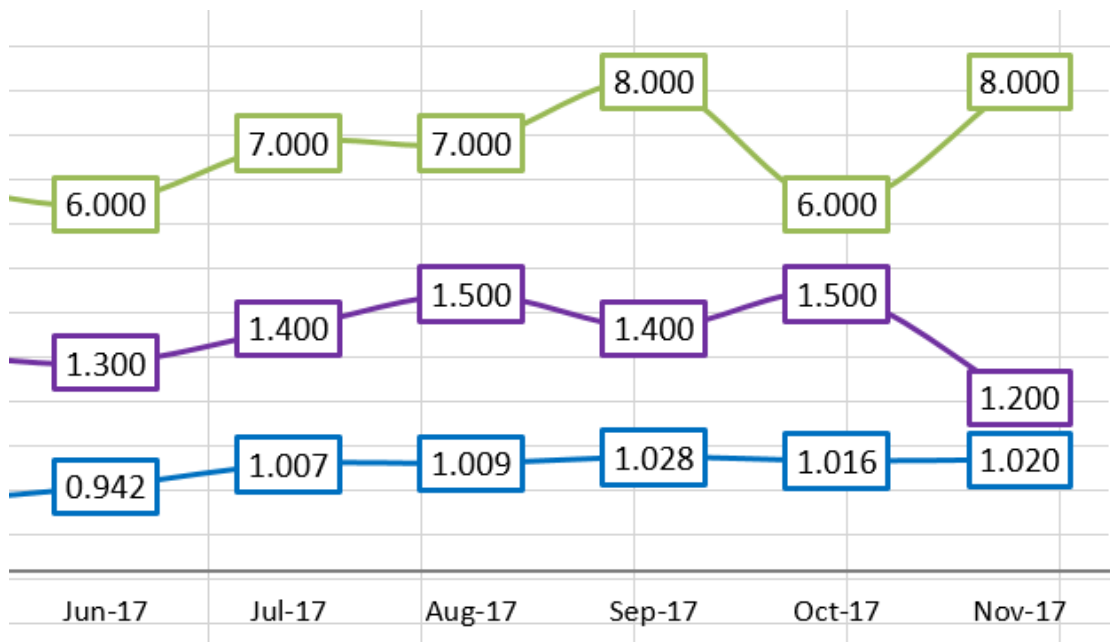
Our Community Link Analysis system provides us with only the groups and their roles, but with no further information about the relations between clients of each community, thus the idea to find which community are representing one family and living under the same roof, the key data here was the Cell_ID tables, along with a hard work on Toad for **feature engineering** (used the ones generated by the CLA such as: betweenness, hub score ..), to finally use the appropriate **network analysis** package from R for the final touch.

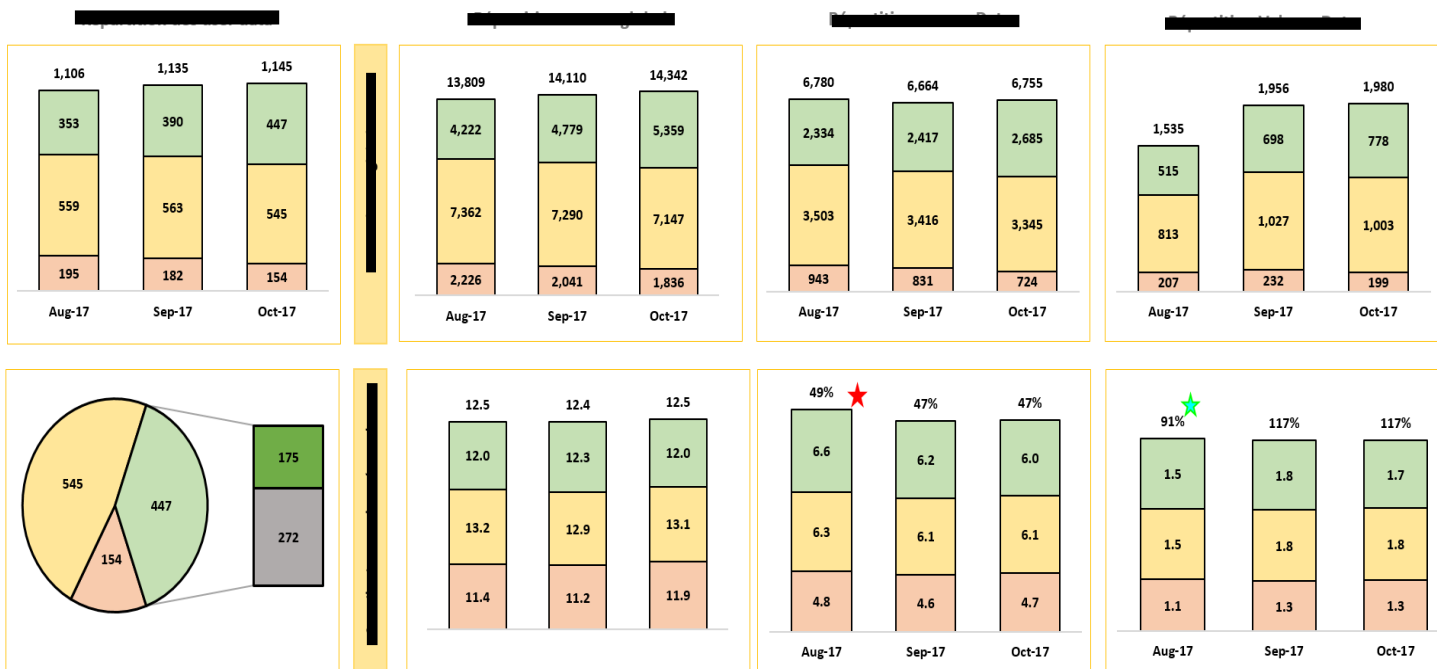
Ad-hoc Analysis / Reports



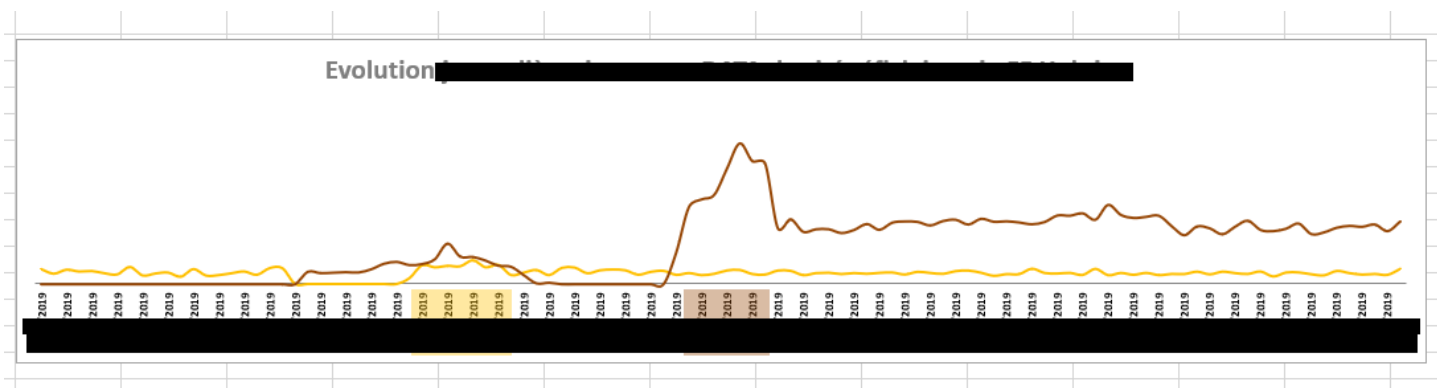
Various demand from different department came to us to solve, to investigate and to find a patterns from our sources in the final goal to make the **data storytelling**.

Understand the **business** behind the demand, create the right SQL request, organize the extracted data with Excel, and at some points go further with PowerPoint for **presentations**.





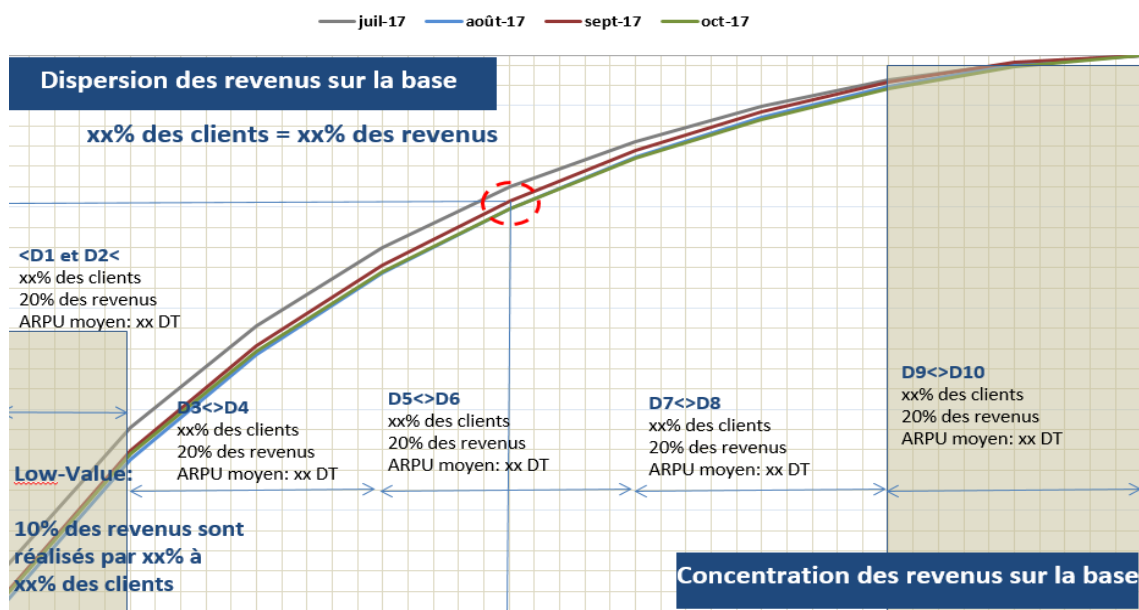
	A	B	C	D	E	F	G	H	I	J
46	>=30	31 533	51 947	39 289	65 338	77 731	104 715	113 038	16 836	500 427
47	Total général									
48										
49										
50	split_actv_mensuel	(Plusieurs éléments)								
51										
52	Somme de	Étiquettes de colonnes								
53	Étiquettes de lignes									
54	inactif	108 835	1 369	201	212	113	185	7 349	#####	
55	<2	51 943	629	86	106	51	88	7 614	60 517	
56	[2,3[	12 719	182	10	28	14	14	2 010	14 977	
57	[3,4[	9 248	175	21	8	2	11	1 584	11 049	
58	[4,5[	7 641	153	14	12	5	12	1 266	9 103	
59	[5,7[	12 521	206	33	16	7	13	2 116	14 912	
60	[7,9[	10 919	159	30	25	16	26	1 785	12 960	
61	>=9	289 108	1 832	292	319	175	305	43 648	#####	
62	Total général								#####	



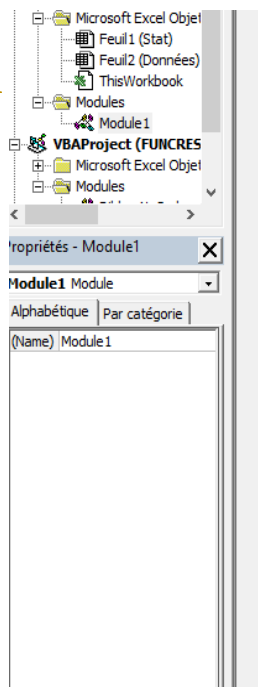
Impact des Campagnes Supplémentaires fin Octobre



- # Users FF - Revenu FF	Un apport supplémentaire total de xx K TND durant les 4 derniers jours d'Octobre (impact sur 7 jours)				
	28/10	29/10	30/10	31/10	7 jours après la fin du campagne
DATA	xx K xx K TND	xx K xx K TND	xx K xx K TND	xx K xx K TND	# : xx K \$: + xx K DT



VBA



```

End Sub

Sub AJR_SLIDE7()
'Copiage ARPU Total
For i = 19 To 24
    Nom_shape = "ARPU_" & CStr(i)
    ActiveSheet.Shapes.Range(Array(Nom_shape)).Select
    Selection.ShapeRange(1).TextFrame2.TextRange.Characters.Text = Format(Range("F" & i).Value, "#,##0.0") & "DT"
Next
'Copiage ARPU et volume Data des user data
For i = 19 To 24
    Nom_shape2 = "ARPUUSER_" & CStr(i)
    ActiveSheet.Shapes.Range(Array(Nom_shape2)).Select
    Selection.ShapeRange(1).TextFrame2.TextRange.Characters.Text = Format(Range("G" & i).Value, "#,##0.0") & "DT" & Chr(13) & Format(Range("H" & i).Value, "#,##0.0")
Next
' Copiage % User DATA
For i = 19 To 24
    Nom_shape3 = "USER_" & CStr(i)
    ActiveSheet.Shapes.Range(Array(Nom_shape3)).Select
    Selection.ShapeRange(1).TextFrame2.TextRange.Characters.Text = Format(Range("D" & i).Value, "0.0,%")
Next
' Copiage Nombre des clients
For i = 19 To 24
    Nom_shape4 = "CLIENT_" & CStr(i)
    ActiveSheet.Shapes.Range(Array(Nom_shape4)).Select
    Selection.ShapeRange(1).TextFrame2.TextRange.Characters.Text = Format(Range("B" & i).Value, "#,##0.0") & "k clients"
Next
' Copiage nombre, valeur et volume par split Revenu

```

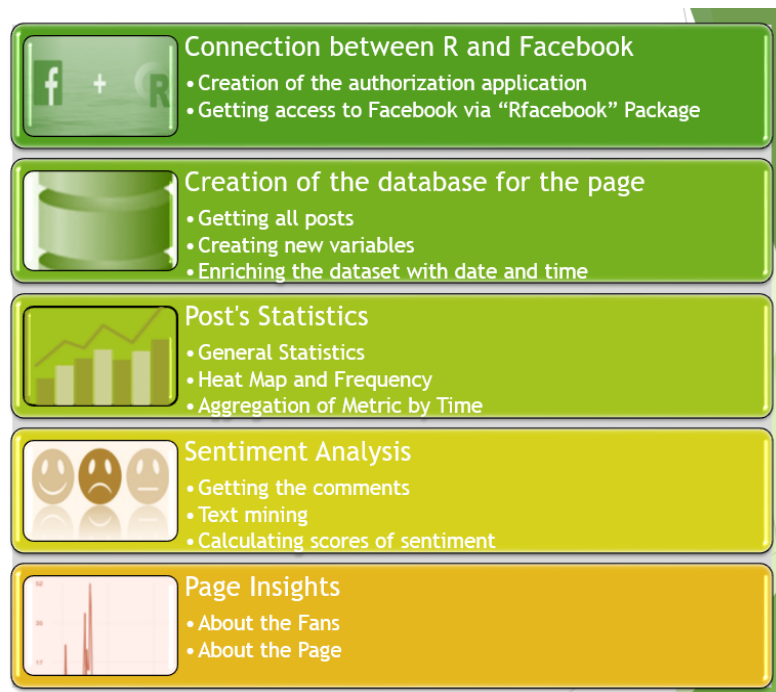
Data Science Projects @ Freelance

I had the opportunity to work on various topics and connect to different teams with the agile methodology.

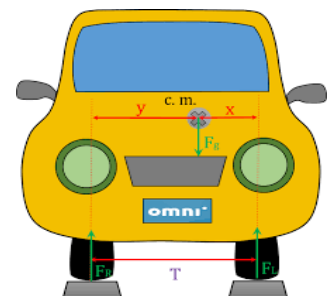
Facebook Page Analysis



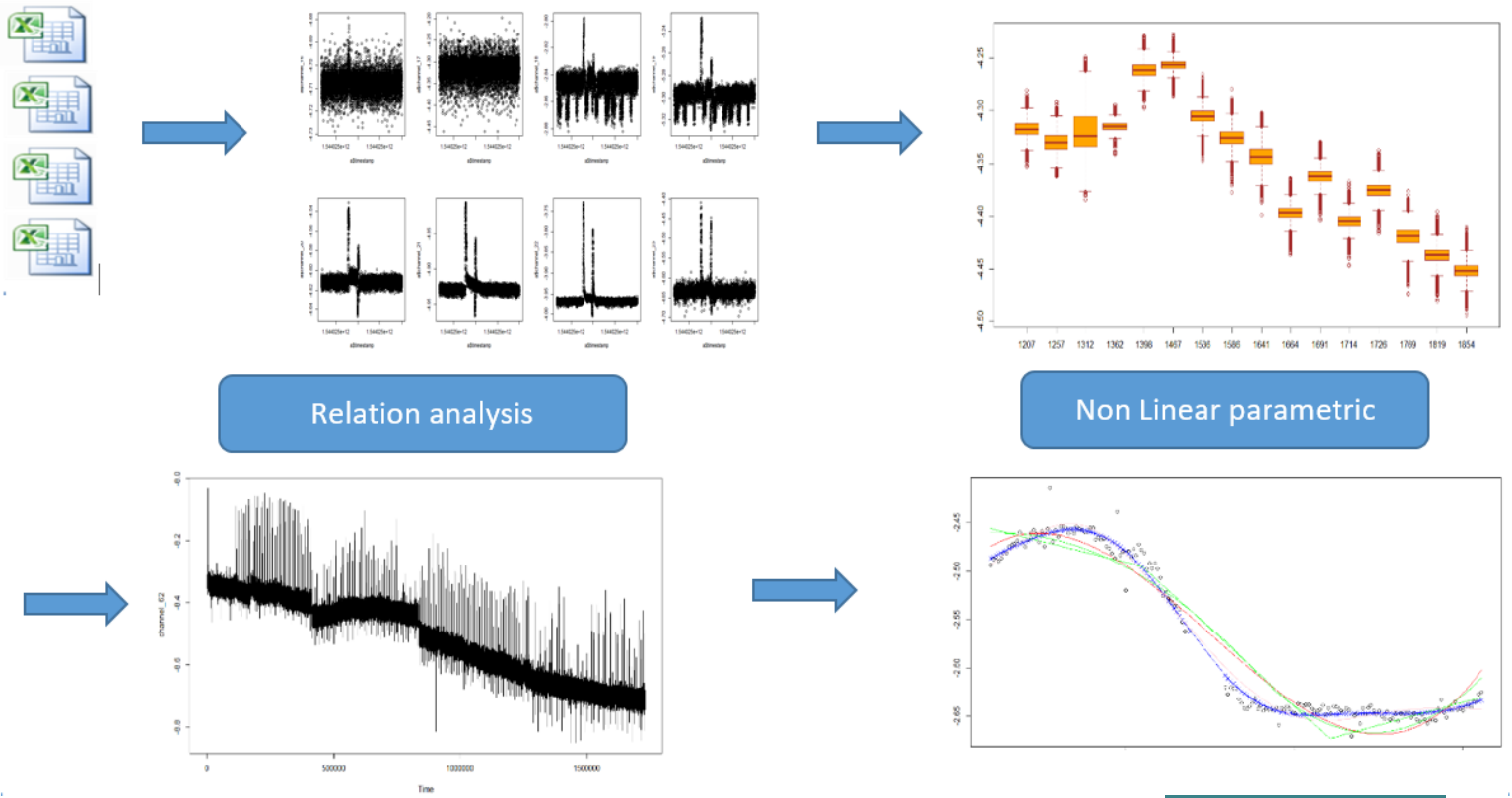
The main idea was to do an overall analysis for one particular page which we have its credentials, to look at its numbers insights and to do some **text mining** in order to get sentiment analysis.



Vehicle Weight Estimation



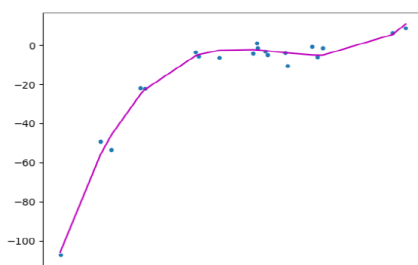
To optimize its logistic line, a company want to predict the mass of vehicles on a connected weighing system with a maximum error less or equal to 5%.



Data Tutor Tutor

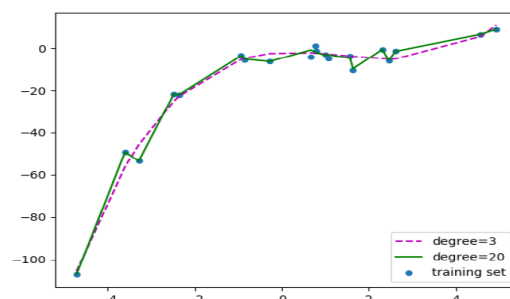
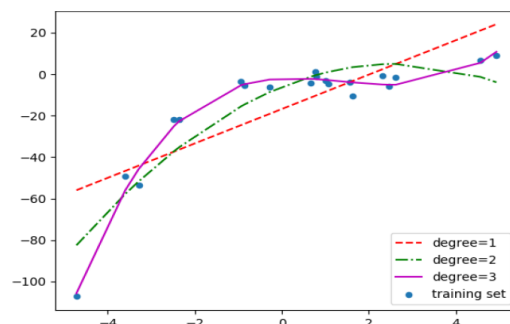
I had my chance to be a **tutor** and introduce a course about Machine Learning: algorithms and examples for a class of 5 persons.

Polynomial Regression



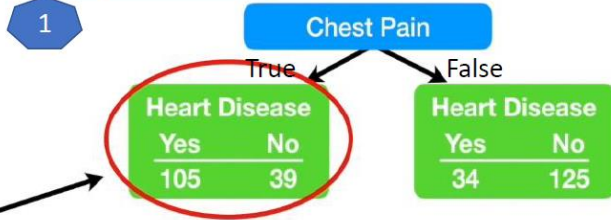
Polynomial of degree = 3
The error (RMSE) continue to decrease

To avoid high degree that leads to overfitting



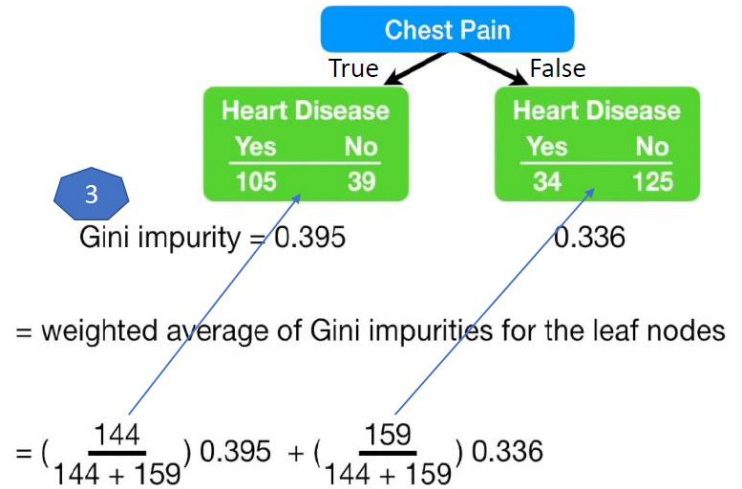
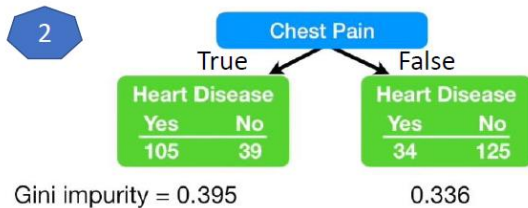
Decision Tree

Steps with example

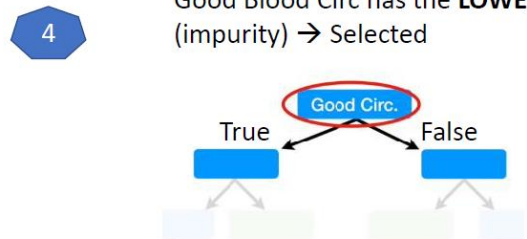


Gini is calculated as follow :

$$= 1 - (\text{the probability of "yes"})^2 - (\text{the probability of "no"})^2$$



Good Blood Circ has the **LOWEST** Gini (impurity) → Selected

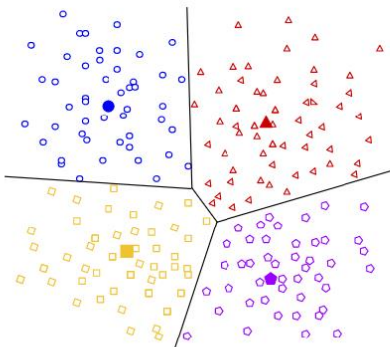


53

Different Types of Clustering Algorithm

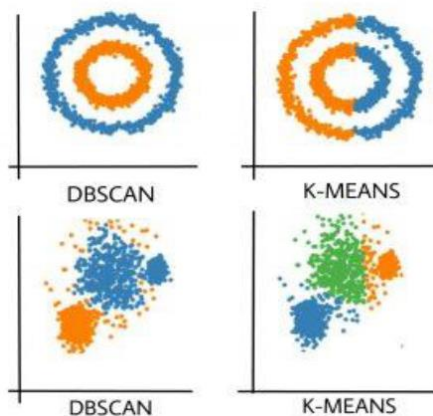
Centroid based method:

- sensitive to initial conditions and outliers
- determine all clusters at once



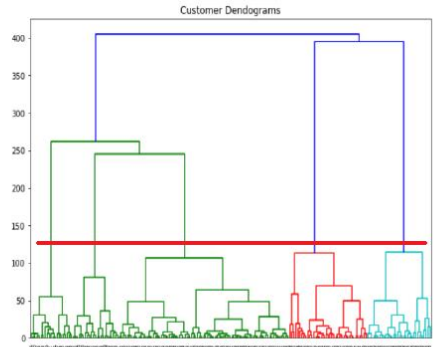
Density based method:

- connects areas of high example density into clusters
- outliers are out of scoop



Hierarchical method:

- creates a tree of clusters
- cutting the tree at the right level

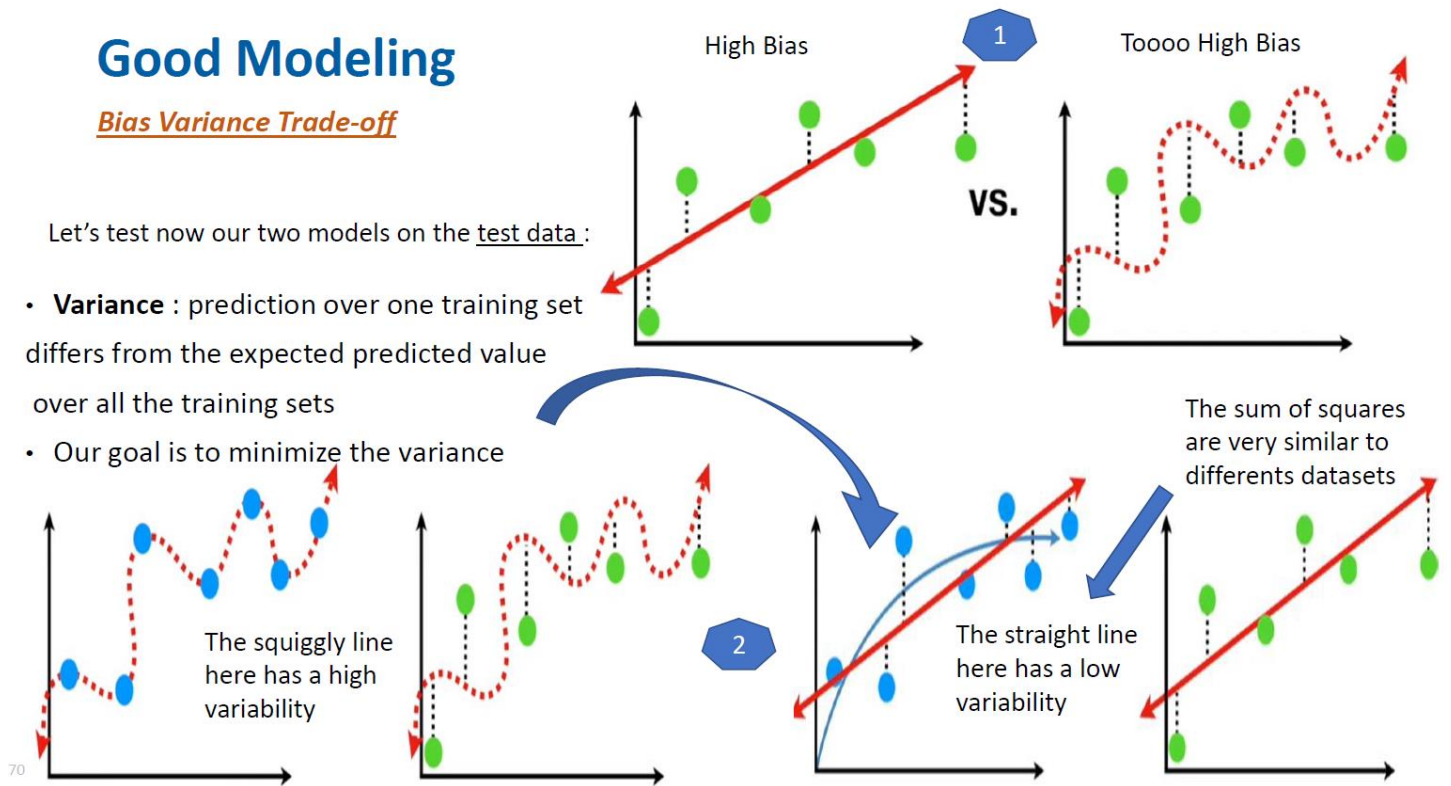


Good Modeling

Bias Variance Trade-off

Let's test now our two models on the test data :

- **Variance** : prediction over one training set differs from the expected predicted value over all the training sets
- Our goal is to minimize the variance



I also had another training session where I animated about the use of machine learning with **IBM Watson** for a class of 20 persons.

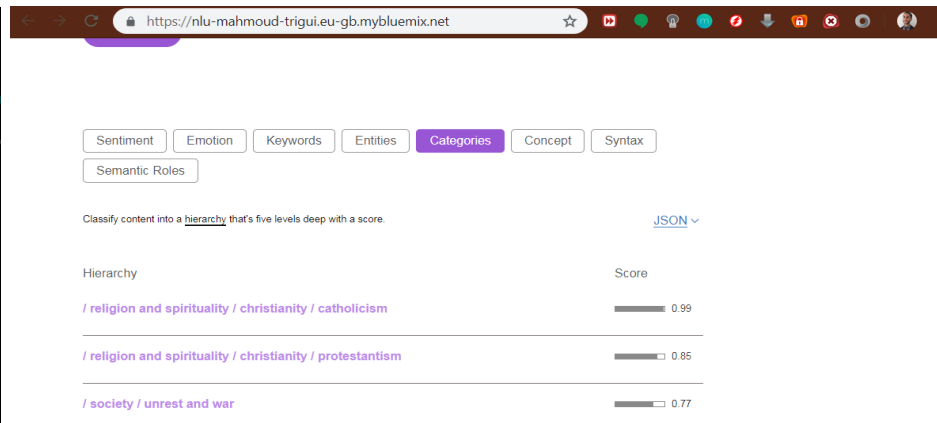
Connecting to CLI

- ▶ Go to the folder of downloaded application, right-click anywhere in order to launch your shell in that location
- ▶ Type « `ibmcloud login` » and enter your email and password
- ▶ Then when asked to select a region, it's preferable to choose the same region as the service was created with (London is in Great-Britain)

```
C:\Users\mt75907\Downloads\Watson-Service-Name-nodejs>ibmcloud app push
Appel de 'cf push'...

Pushing from manifest to org trigui.mahmoud7@gmail.com / space dev as trigui.mahmoud7@gmail.com...
Using manifest file C:\Users\mt75907\Downloads\Watson-Service-Name-nodejs\manifest.yml
Getting app info...
Creating app with these attributes...
+ name: Mahmoud-Trigui-App
+ path: C:\Users\mt75907\Downloads\Watson-Service-Name-nodejs
+ command: npm start
+ memory: 128M
+ services:
+ Watson-Service-Name
+ routes:
+ mahmoud-trigui-app.eu-gb.mybluemix.net

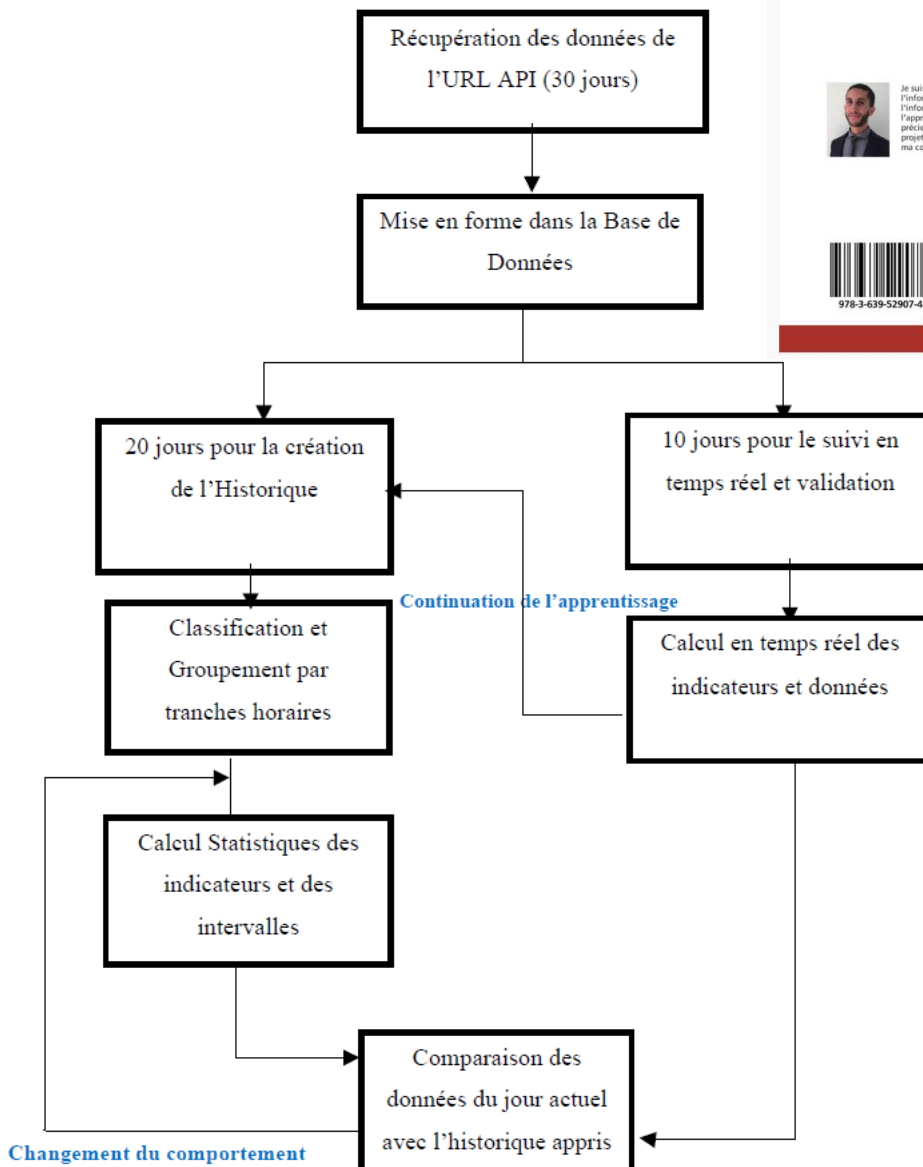
Creating app Mahmoud-Trigui-App...
Mapping routes...
Binding services...
```



Alarm Automatic management



The idea is to develop an automatic model to know from the owner's habits, if it's the owner who tries to enter or it's an intrusion based on the owner behavior, and then decide to launch the alarm or to deactivate it automatically.



```

private static final int BUFFER_SIZE = 4096;
private static String baseOutputFolder = "C:\\Users\\LAAS\\Desktop\\Recupjava\\AprèsRéunion";
private static String baseUrl = "https://bo-api...";
private static List<String> clientsIds = Arrays.asList("6C0B84...");
private static LocalDate startDate = LocalDate.of(2015, 10, 01);
private static LocalDate endDate = LocalDate.of(2015, 10, 21);

public static void main(String[] args) throws IOException
{
    for (String clientId : clientsIds)
    {
        System.out.println("CLIENT " + clientId);
        String clientUrl = baseUrl + "/" + clientId + "/";
        File outputFolder = new File(baseOutputFolder, clientId);
        outputFolder.mkdirs();
        LocalDate currentDate = startDate;
        while (!currentDate.isAfter(endDate))
        {
            URL url = new URL(clientUrl + currentDate.toString());
        }
    }
}

```

Run Connexbd

```

"C:\Program ...
CLIENT 6C0B841022DB
Recup de la ressource : https://bo-api. /2015-10-01
Recup de la ressource : https://bo-api. /2015-10-02
Recup de la ressource : https://bo-api. /2015-10-03
Recup de la ressource : https://bo-api. /2015-10-04

```

La fin moyenne de la 1 ère classe est 8 : 18 h
 avec un écart-type de 2 minutes
 L'intervalle pour la 1 ère classe est [8 : 16 ; 8 : 20]
 La fin de la 1 ère tranche horaire est 8 : 30 h

La fin moyenne de la 2 ème classe est 17 : 43 h
 avec un écart-type de 16 minutes
 L'intervalle pour la 2 ère classe est [17 : 27 ; 17 : 59]
 La fin de la 2 ème tranche horaire est 18 : 0 h

Figure 10. Exemple de début et fin des tranches horaires

	Date_Heure	Devtype	Numero	Fonction	Fonction_nom	Sensib_Angle	Etat_porte	Ouv_fer	Ang_deb	Ang_fin
40	2015-09-03 07:39:55	IntelliTag	4 6		Presence	-52	416	-2188	1743	149
41	2015-09-03 08:10:37	IntelliTag	3 32		AlarmeAcc	4	4	2	1	10
42	2015-09-03 08:11:15	IntelliTag	3 32		AlarmeAcc	5	4	1	5	0
43	2015-09-03 08:17:44	Remote	2	armed.userRemote	Activ	NA	NA	NA	NA	NA
44	2015-09-03 08:17:51	IntelliTag	5 32		AlarmeAcc	8	1	2	0	26
45	2015-09-03 08:17:52	IntelliTag	5 30		AlarmeMag	29	9	826	-709	-253
46	2015-09-03 08:17:56	IntelliTag	5 32		AlarmeAcc	10	4	1	15	2
47	2015-09-03 08:17:58	IntelliTag	5 30		AlarmeMag	3	4	506	-254	-333
48	2015-09-03 08:18:09	IntelliTag	5 30		AlarmeMag	0	4	493	-276	-321
49	2015-09-03 08:19:58	Remote	2	Check-out	5	0	0	0	0	0
50	2015-09-03 08:19:59	Remote	2	IsLost=1		0	0	0	0	0

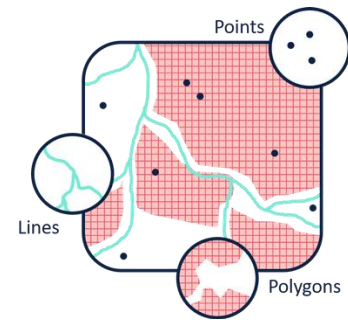
Figure 14. Activation de l'alarme avec information préalable

Machine Learning Competitor

I have been competing for over 18 months & over 14 challenges and I'm not intending to stop any way for a simple reason: I just keep learning new things every day, new algorithm, new way of thinking, new approach, new trick, new package, new optimization.

I gained so much knowledge through this period, because you need to be always up to date to latest news for the data science community or you will find yourself left behind and just repeating what you have gained!

Spatial Data



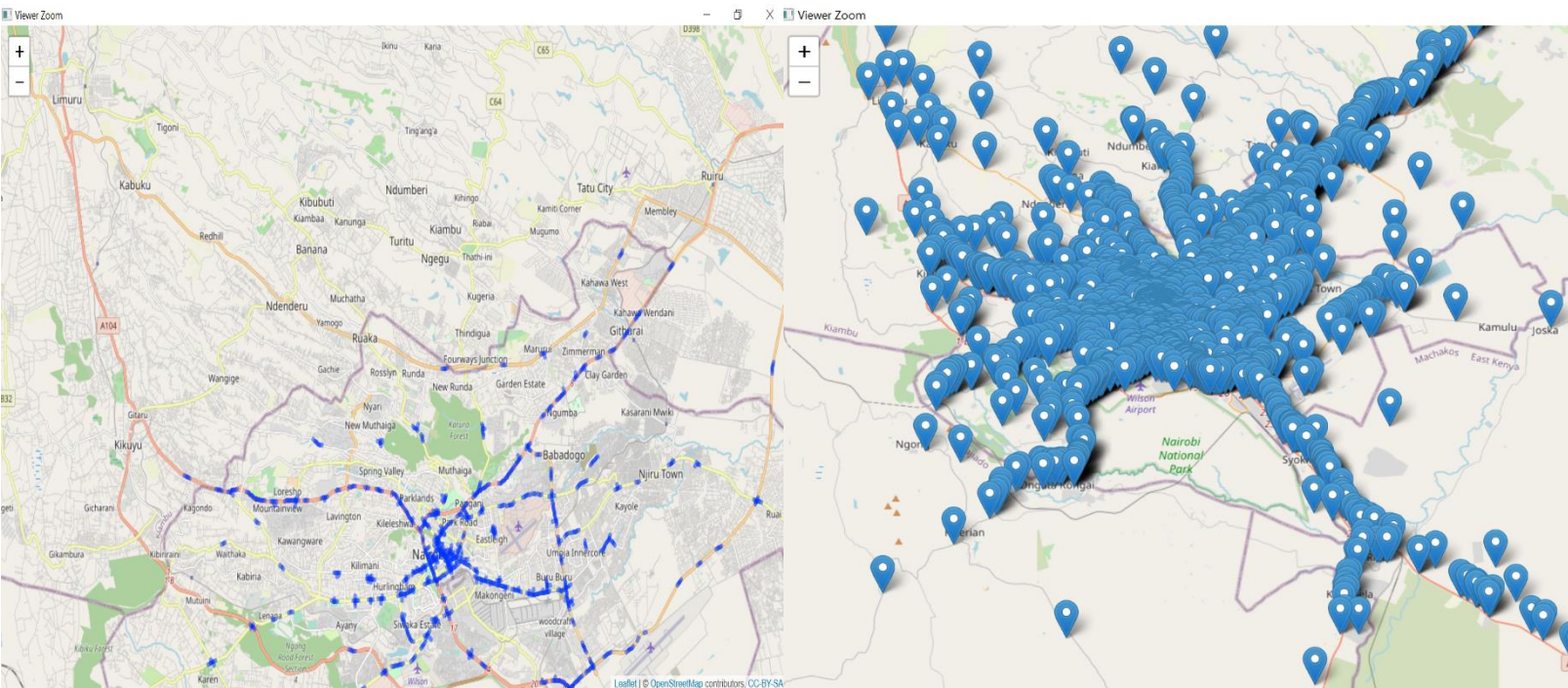
OSM & Roads

```
91 node=as.data.frame(read.csv("movement-junctions-to-osm-nodes-nairobi-2019.csv"))
92 node$junction_id=NULL
93 uber=as.data.frame(read.csv("movement-speeds-quarterly-by-hod-nairobi-2019-Q2.csv"))
94
95 uber <- uber %>% select(-c(1,2,4:6))
96 node$lon=0
97 node$lat=0
98 for (i in 1:nrow(node)) {
99   tryCatch({
100     node$lon[i] <- get_osm(node(node$osm_node_id[i]), source = src)$nodes$attrs$lon
101     node$lat[i] <- get_osm(node(node$osm_node_id[i]), source = src)$nodes$attrs$lat
102   }, error=function(e){})
103 }
104
105 for (i in 1:nrow(df)) {
106   tryCatch({
107     coordinates <- list(c(as.numeric(df[i,25]), as.numeric(df[i,24])), c(as.numeric(df[i,27]),
108     x <- ors_directions(coordinates, profile = "cycling-regular",elevation = TRUE,extra_info =
109     df$dist[i]=x$features[[1]]$properties$summary$distance
110     df$dur[i]=x$features[[1]]$properties$summary$duration
111     df$ascent[i]=x$features[[1]]$properties$segments[[1]]$ascent
112     df$descent[i]=x$features[[1]]$properties$segments[[1]]$descent
113     df$step[i]=x$features[[1]]$properties$extras$steepness$summary[[1]]$amount
114     l[[i]]=x
115   }, error=function(e){})
116 }
```

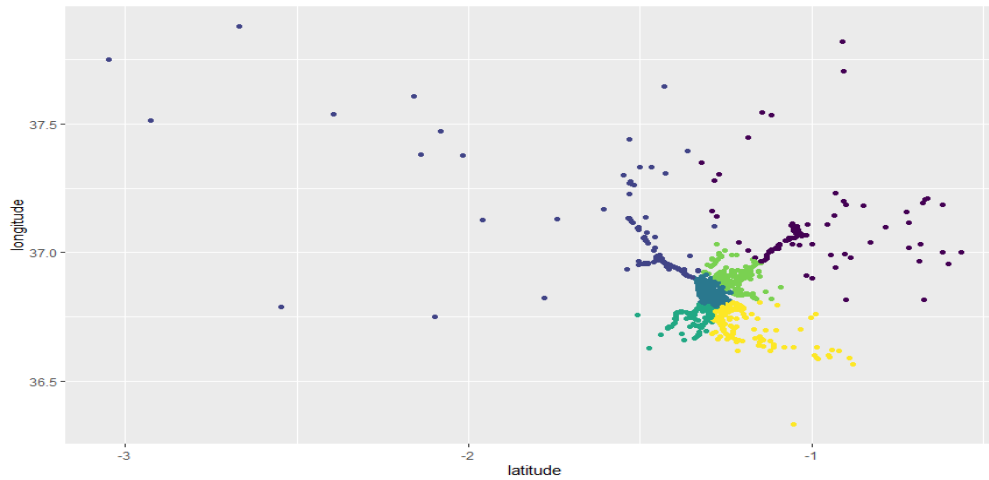
Shape & Raster files, Projections

```
70 #####spatial#####
71
72 cr = "+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0"
73 #convert X,Y to shapefile
74 train_sp <- st_as_sf(train, coords = c("X", "Y"), crs = cr)
75 #train_sp_ <- SpatialPointsDataFrame(coords = c("X", "Y"), data = train, proj4string = CRS(
76 #mapview(train_sp)
77 train_sp_ <- as(train_sp, "Spatial") #SpatialPointsDataFrame
78 coor <- train[,c(2,3)]
79 proj4string(coor) <- crs(lakes_shp)
80 coor <- spTransform(coor, CRS(cr))
81
82 d2 <- data.frame(lon=df$X,lat = df$Y,elevation = df$elevation)
83 rast = raster::rasterFromXYZ(d2)
84 raster::crs(rast) = cr
85 d3 <- data.frame(lon=df$X,lat = df$Y, target = df$target_2015)
86 rast_m = rasterFromXYZ(d3)
87 b1 <- brick(rast, rast_m)
88
89 terr <- raster::terrain(rast,opt = c('slope','aspect','tpi','tri','roughness','flowdir'),
90 unit = 'degrees',neighbors = 8)
91 raster::crs(rast_m) = c("+proj=longlat +datum=WGS84 +no_defs +ellps=WGS84 +towgs84=0,0,0")
92
93 #----External Data----
94 utmStr <- "+proj=utm +zone=%d +datum=NAD83 +units=m +no_defs +ellps=GRS80 +towgs84=0,0,0"
95 crs <- CRS(sprintf(utmStr, 32))
96 ptsUTM <- spTransform(coor, crs)
97 n <- length(ptsUTM)
98 ptsUTM_data = ptsUTM
99
100 water_sf <- sf::read_sf("MWI_water_areas_dcw.shp")
101 water_shp <- readOGR(".", "MWI_water_areas_dcw")
102 pUTM <- spTransform(water_shp, crs)
103 a <- apply(gDistance(ptsUTM, pUTM,byid=TRUE),2,min)
104 #a <- apply(gDistance(coor, water_shp,byid=TRUE),2,min)
105 water_dist <- as.vector(a)
106 ptsUTM_data$water_dist = water_dist
```

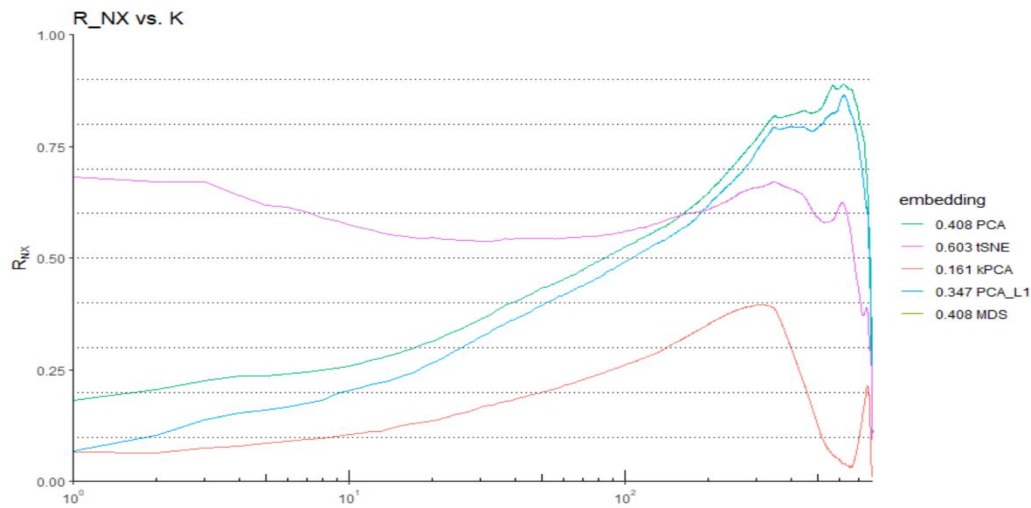
Geospace Analysis



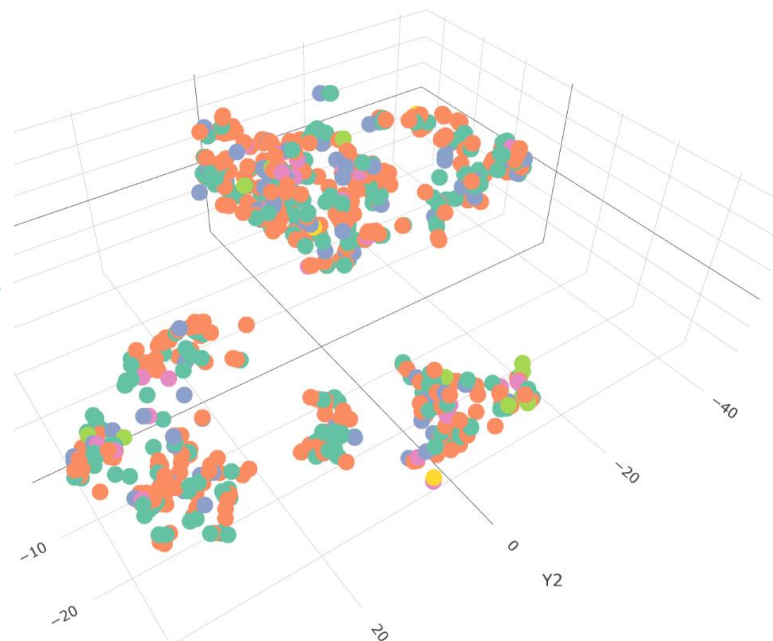
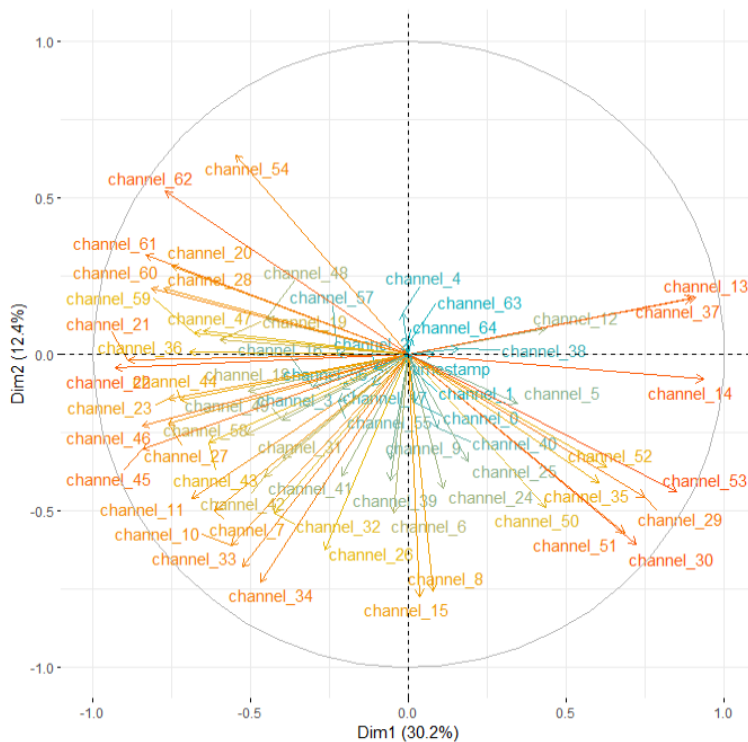
Various Clustering



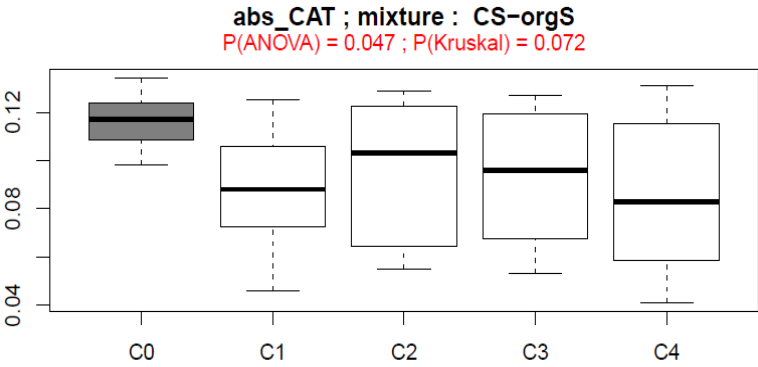
Dimensionality Reduction



Variables - PCA



ANOVA Analysis

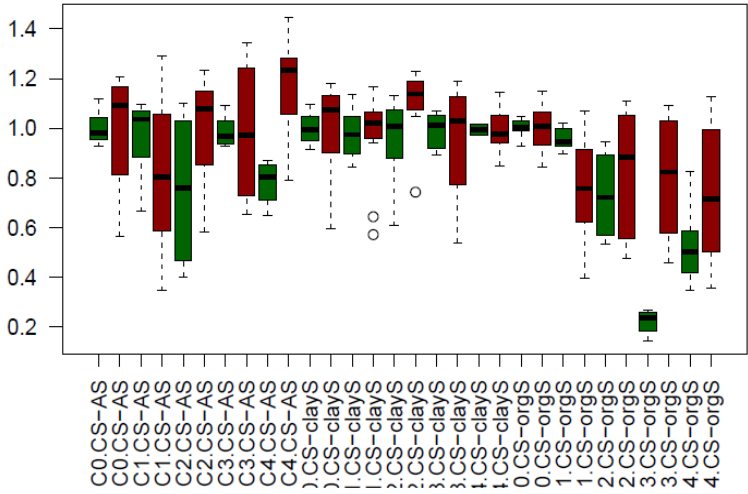


Tukey post hoc test

	diff	lwr	upr	p adj
C1-C0	-0.0275833	-0.0575680	0.0024013	0.0853935
C2-C0	-0.0207500	-0.0507347	0.0092347	0.3031383
C3-C0	-0.0231667	-0.0531513	0.0068180	0.2029738
C4-C0	-0.0307500	-0.0607347	-0.0007653	0.0417953
C2-C1	0.0068333	-0.0231513	0.0368180	0.9672798
C3-C1	0.0044167	-0.0255680	0.0344013	0.9935809
C4-C1	-0.0031667	-0.0331513	0.0268180	0.9982319
C3-C2	-0.0024167	-0.0324013	0.0275680	0.9993891
C4-C2	-0.0100000	-0.0399847	0.0199847	0.8797444
C4-C3	-0.0075833	-0.0375680	0.0224013	0.9525919

Pairwise t-test as post hoc test

	C0	C1	C2	C3
C1	0.0606008	NA	NA	NA
C2	0.1401763	0.7472369	NA	NA
C3	0.1121024	0.8210262	0.8210262	NA
C4	0.0546853	0.8210262	0.7020563	0.7472369



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
v_days	1	0.2449093	0.2449093	4.957332	0.0269371
v_conc_1	4	0.5722125	0.1430531	2.895610	0.0228805
v_soil_type	2	2.0323626	1.0161813	20.569035	0.0000000
Residuals	233	11.5110040	0.0494035	NA	NA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
lays	1	0.2449093	0.2449093	6.1591007	0.0138545
onc_1	4	0.5722125	0.1430531	3.5975706	0.0073390
oil_type	2	2.0323626	1.0161813	25.5554304	0.0000000
lays:v_conc_1	4	0.8338934	0.2084734	5.2427912	0.0004800
lays:v_soil_type	2	0.0430311	0.0215156	0.5410839	0.5829225
onc_1:v_soil_type	8	1.2535646	0.1566956	3.9406581	0.0002300
lays:v_conc_1:v_soil_type	8	0.9903509	0.1237939	3.1132293	0.0024073
iduals	211	8.3901640	0.0397638	NA	NA

Assessments, Courses & Skills

Pluralsight

M Mahmoud Trigui

 Implementing the Data Science Workflow in... PROFICIENT 186 73rd percentile VERIFIED 2.11.2021	 Business Communications with Data PROFICIENT 195 78th percentile VERIFIED 2.11.2021	 Microsoft Excel Data Insights EXPERT 214 86th percentile VERIFIED 2.10.2021	 Data Wrangling with R EXPERT 246 95th percentile VERIFIED 2.10.2021	 Cleaning Data with R EXPERT 227 91st percentile VERIFIED 2.9.2021
 Bringing Data Science into the Business EXPERT 223 89th percentile VERIFIED 2.9.2021	 Feature Engineering EXPERT 202 81st percentile VERIFIED 2.9.2021	 Programming R for Data Analysis EXPERT 207 83rd percentile VERIFIED 2.3.2021	 Interpreting Data with R EXPERT 203 82nd percentile VERIFIED 10.2.2020	 Programming Data Structures in R EXPERT 221 89th percentile VERIFIED 10.2.2020
 R as a Language EXPERT 232 92nd percentile VERIFIED 10.1.2020	 Data Science Literacy EXPERT 230 91st percentile VERIFIED 10.1.2020	 Creating and Using R Functions EXPERT 227 91st percentile VERIFIED 9.30.2020	 Machine Learning Literacy EXPERT 216 87th percentile VERIFIED 9.30.2020	 Data Analytics Literacy EXPERT 244 95th percentile VERIFIED 4.7.2020

Skill Assessments

LinkedIn

Badges (3)

To retake (1)



Machine Learning
Badge earned 8/21/2020



R (Programming Language)
Badge earned 9/26/2019



Transact-SQL (T-SQL)
Badge earned 3/26/2020

Acclaim

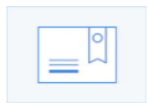
 Data Science Orientation Coursera	 Data Science Foundations - Level 1 IBM	 SAS Programming 1: Essentials SAS
 SAS Programming 1: Essentials SAS	 Data Science Foundations - Level 2 (V2) IBM	 Build Your Own Chatbot - Level 1 IBM
 Statistics 101 IBM	 Watson Assistant Foundations IBM	 Node-RED: basics to bots IBM
 Statistics 1: Introduction to ANOVA, Regression, and Logistic... SAS	 Machine Learning with R - Level 1 IBM	 Watson Studio Essentials IBM
 Exploratory Data Analysis SAS	 Introduction to Statistical Concepts SAS	 Quality Methods SAS
 Decision Making with Data SAS	 Correlation and Regression SAS	 Predictive Modeling and Text Mining SAS
 IBM Watson Machine Learning Essentials IBM	 Data Refinery Essentials IBM	 Applied Data Science with R - Level 2 IBM

DataCamp

 Understanding and Interpreting Data ADVANCED Score: 156 Percentile: 97% View Results Retake	 Statistics Fundamentals with R ADVANCED Score: 163 Percentile: 98% View Results Retake	 Machine Learning Fundamentals in R ADVANCED Score: 164 Percentile: 98% View Results Retake
 R Programming ADVANCED Score: 138 Percentile: 90% View Results Retake	 Importing & Cleaning Data with R ADVANCED Score: 144 Percentile: 93% View Results Retake	 Data Manipulation with R ADVANCED Score: 144 Percentile: 93% View Results Retake

Coursera

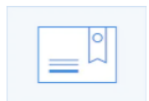
My Courses



Machine Learning for Data Analysis

Wesleyan University
Grade Achieved: 100%

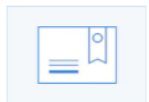
[Add to LinkedIn](#) ▼



Text Mining and Analytics

University of Illinois at Urbana-Champaign
Grade Achieved: 88%

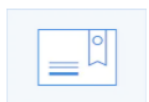
[Add to LinkedIn](#) ▼



Inferential and Predictive Statistics for Business

University of Illinois at Urbana-Champaign
Grade Achieved: 88%

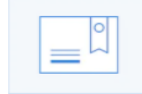
[Add to LinkedIn](#) ▼



Machine Learning

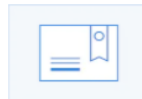
Stanford University
Grade Achieved: 94%

[Add to LinkedIn](#) ▼



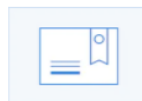
Managing Big Data with MySQL

Duke University
Grade Achieved: 93%



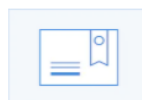
What is Data Science?

IBM
Grade Achieved: 92%



Cluster Analysis in Data Mining

University of Illinois at Urbana-Champaign
Grade Achieved: 85%



Regression Models

Johns Hopkins University
Grade Achieved: 100%

My Certificates

👁 Everyone on edX

[Edit](#)

edX

Verified Certificate

Introduction to Python for Data Science

From
Microsoft

Completed on 4/26/2016

[View Certificate](#) ↗

Verified Certificate

Querying with Transact-SQL

From
Microsoft

Completed on 5/4/2016

[View Certificate](#) ↗

Verified Certificate

Analyzing and Visualizing Data with Excel

From
Microsoft

Completed on 5/17/2016

[View Certificate](#) ↗

Certificate

Data Science: Data to Insights

From
MITProfessionalX

Completed on 11/22/2016

[View Certificate](#) ↗

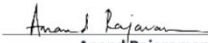
Stanford | ONLINE
STATEMENT OF ACCOMPLISHMENT

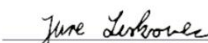
October 27, 2018

Mahmoud Trigui
has successfully completed a free online offering of

Mining Massive Datasets

This course covers "big-data" algorithms, including locality-sensitive hashing, PageRank, stream algorithms, clustering, social-network graph analysis, large-scale machine learning, recommendation systems, computational advertising, and dimensionality reduction. The SoA is awarded to those scoring at least 50% of the marks, half of which are based on homeworks and half on a timed final exam.


Anand Rajaram
Founding Partner
Rocketship.vc


Jure Leskovec
Associate Professor of Computer Science
Stanford University


Jeffrey D. Ullman
S. W. Ascherman Professor of Engineering
(emeritus)
Stanford University

PLEASE NOTE: SOME ONLINE COURSES MAY DRAW ON MATERIAL FROM COURSES TAUGHT ON-CAMPUS BUT THEY ARE NOT EQUIVALENT TO ON-CAMPUS COURSES. THIS STATEMENT DOES NOT AFFIRM THAT THIS PARTICIPANT WAS ENROLLED AS A STUDENT AT STANFORD UNIVERSITY IN ANY WAY. IT DOES NOT CONFER A STANFORD UNIVERSITY GRADE, COURSE CREDIT OR DEGREE, AND IT DOES NOT VERIFY THE IDENTITY OF THE PARTICIPANT.

Authenticity can be verified at <https://verify.lagunita.stanford.edu/50A/0d5359fd79a642d2ad742ba9c4dd802b>

Stanford | ONLINE
STATEMENT OF ACCOMPLISHMENT

March 3, 2020

Mahmoud Trigui
has successfully completed a free online offering of

Statistical Learning

In order to receive a statement of accomplishment, participants were required to score 50 or more of the 100 points attainable on the online quizzes.


Trevor Hastie
John A Overdeck Professor of Statistics
Stanford University


Rob Tibshirani
Professor in Health Research and Policy and Statistics
Stanford University

PLEASE NOTE: SOME ONLINE COURSES MAY DRAW ON MATERIAL FROM COURSES TAUGHT ON-CAMPUS BUT THEY ARE NOT EQUIVALENT TO ON-CAMPUS COURSES. THIS STATEMENT DOES NOT AFFIRM THAT THIS PARTICIPANT WAS ENROLLED AS A STUDENT AT STANFORD UNIVERSITY IN ANY WAY. IT DOES NOT CONFER A STANFORD UNIVERSITY GRADE, COURSE CREDIT OR DEGREE, AND IT DOES NOT VERIFY THE IDENTITY OF THE PARTICIPANT.

Authenticity can be verified at <https://verify.lagunita.stanford.edu/50A/59d49a34a8ad4b019e47ce2c473501b9>

Stanford | ONLINE
STATEMENT OF ACCOMPLISHMENT

Mahmoud Trigui

has successfully completed a free online offering of

SQL

with **Statement of Accomplishment.**

MIT ProX, Udacity & DataQuest



Massachusetts Institute of Technology

This is to certify that

Mahmoud Trigui



has successfully completed

Data Science: Data to Insights

October 4 – November 15, 2016

Bhaskar Pant
Executive Director
MIT Professional Education

Devavrat Shah
Professor & Director of Statistics and Data Science Center
MIT, EECS, IDSS, LIDS, Statistics and Data Science Center

Philippe Rigollet
Associate Professor
MIT, Department of Mathematics, IDSS, LIDS, Statistics and Data Science Center and Broad Institute

Verified Certificate of Completion

Apr 2018



Mahmoud Trigui

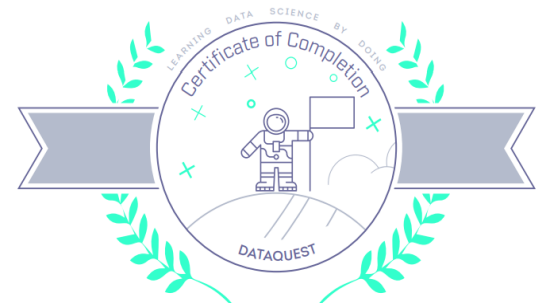
has successfully completed the

مبادرة مليون مبرمج عربي </10^6>
1 MILLION ARAB CODERS INITIATIVE

Data Analyst Track

1 Million Arab Coders Initiative

Sebastian Thrun
Founder, Udacity



Mahmoud Trigui

Successfully completed the following path

DATA ANALYST IN R

Issued on 2020-07-26

Verify this certificate at https://app.dataquest.io/verify_cert/EKX8Z52UXNJHEBBPRYIH/

English Certificate

EF SET

English Certificate

Mahmoud Trigui

has successfully completed the EF SET Certificate
and has earned level:



Reading



Listening

Awarded on:
09 Feb 2023



English

Reading

C2 Proficient

You can read virtually any kind of text (factual, literary, technical) and accurately recognize and categorize style and tone. You can understand complex technical writing on unfamiliar subjects on a wide range of topics.

- Can read with ease virtually all forms of written language, including abstract, structurally or linguistically complex texts such as manuals, specialized articles and literary works
- Can understand a wide range of long and complex texts, appreciating subtle distinctions of style, and implicit meaning
- Can understand in detail complex texts likely to be encountered in social, professional or academic life, identifying finer points of detail including attitudes and implied opinions



English

Listening

C2 Proficient

You are comfortable in all situations, live or broadcast, that require full comprehension; you are almost never confused or searching for the meaning of words and phrases. You understand nuances of expression and tone, humor and emphasis in all live theatrical presentations, films or broadcast presentations.

- Can understand with ease any kind of spoken language, live or broadcast, even when delivered at fast native speed, provided with time to get familiar with any regional or other accent
- Can understand lectures and presentations with a high degree of colloquialism, regional usage and unfamiliar terminology
- Can follow complex interactions in group discussions and debates, even on abstract, complex and unfamiliar topics

+ Personality



⚙️ Customize

Hi, Mahmoud!

TYPE: Assertive Advocate ?

CODE: INFJ-A ?

ROLE: Diplomat ?

STRATEGY: Confident Individualism ?

Mind



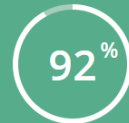
INTROVERTED

Energy



INTUITIVE

Nature



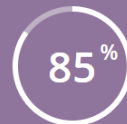
FEELING

Tactics



JUDGING

Identity



ASSERTIVE