DANIELA OSPINA TORO - Data Scientist (Operations Reseach)

1. **PROJECTS (from more recent to less recent).**

Most finished projects have one or two scientific papers published and this information can be verified looking for my full name on Scholar Google.

1) Spaced Repetition Algorithm for Language Learning (2022 - prestudy phase): The objective is to study an open-souce model to adapt to this new application that is being developed.
Freelance data scientist. I am working alone on this.
Software: Python
Techniques: Generation of synthetic data, half-life regression, likelihood-based comparison

$$\underset{\boldsymbol{u}(t_0,t_f]}{\text{minimize}} \ \mathbb{E}\left[\phi(\boldsymbol{m}(t_f),\boldsymbol{n}(t_f)) + \int_{t_0}^{t_f} \ell(\boldsymbol{m}(\tau),\boldsymbol{n}(\tau),\boldsymbol{u}(\tau))d\tau\right]$$
$$\text{subject to } \boldsymbol{u}(t) \geq 0 \ \forall t \in (t_0,t_f), \qquad [4]$$

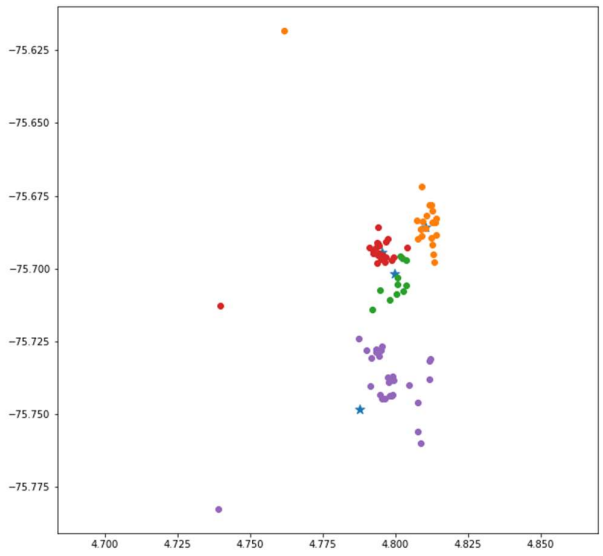
where $\boldsymbol{u}(t_0,t_f]$ denotes the item reviewing intensities from t_0 to t_f, the expectation is taken over all possible realizations of the associated counting processes and (item) recalls, the loss function is nonincreasing (nondecreasing) with respect to the recall probabilities (forgetting rates and intensities) so that it rewards long-lasting learning while limiting the number of item reviews, and $\phi(\boldsymbol{m}(t_f),\boldsymbol{n}(t_f))$ is an arbitrary penalty function. [The penalty function $\phi(\boldsymbol{m}(t_f),\boldsymbol{n}(t_f))$ is necessary to derive the optimal reviewing intensities $\boldsymbol{u}^*(t)$.] Here, note that the forgetting rates $\boldsymbol{n}(t)$ and recall probabilities $\boldsymbol{m}(t)$, as defined by Eqs. **2** and **3**, depend on the reviewing intensities $\boldsymbol{u}(t)$ we aim to optimize since $\mathbb{E}[d\boldsymbol{N}(t)] = \boldsymbol{u}(t)dt$.

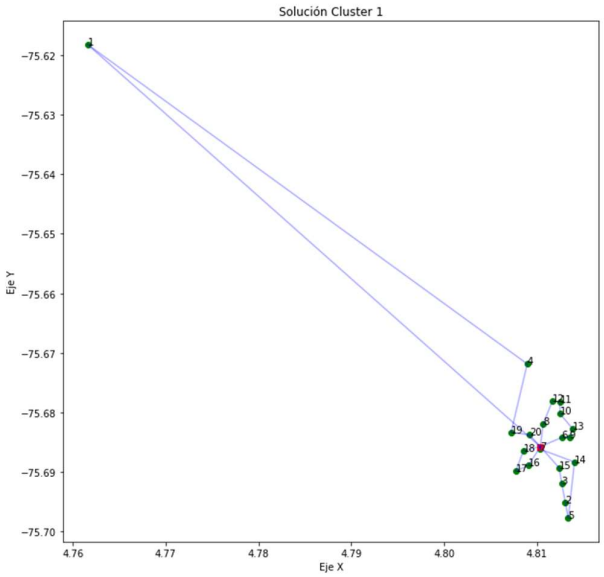
2) **Real Estate Indicators (2022 – in process):** Just starting with this project for a real estate company. It's a web where they will be able to analyze the indicators of an area given a ZCTA5 code.
Freelance data scientist. I am working alone on this.
Software:
R (reticulate, googleway, censusapi), Python (requests), Census API, Shiny framework.
Techniques:
JSON queries reading, data visualization of indicators, regressions, classifiers.

Embed Street View Sphere https://developers.google.com/streetview/web
Code in react? e

Source	Desired Data	Table	Desc	Breadth	Dev Info
census	Percent below poverty level	S1701		5 years zcta, national, graph	https://www.census.gov/data/deve
	median income	S2503	FINANCIAL CHARACTERISTICS	5 years zcta, national, graph	
	population growth	DP05		5 years zcta, national, graph	
	Unemployment rate	S2301	Employment Status - population 20-64	5 years zcta, national, graph	
	selected economic characteristics - job diversity data	DP03		5 years zcta, national, graph	No industry greater than 25% exce
	Median Age	S0101	AGE AND SEX	5 years zcta, national, graph	
	Vacancy Rate	DP04		5 years zcta, national, graph	
	Median Rent	DP04	gross rent	5 years zcta, national, graph	
	5+ units permitted			5 years zcta, national	https://www.census.gov/constructi

3) **Design of recollection routes of usable solid residues to classification and utilization stations (ECA, in Spanish) in Pereira, Colombia using optimization techniques (2022 - about to be completed)**: A project where information about residues collection is estimated starting from official data available from city census. I designed the whole project, chose the technologies to use, trained the trainees, developed the first stages for preprocessing the data. **Project Leader, trainer and developer. The project included also two trainees.**
Software: Python (sklearn, matplotlib, geopy, numpy, pandas), Doplex.mp.model (IBM), Google Maps API
Techniques: K–means clusterization, MDVRP with demand windows, CVRP with demand windows.





4) Solution to the Problem of Localization, Routing and Inventory in Multiobjective Environments (2021 - still going on).

Coresearcher, trainer.

The project is our first approach to the IRP and it includes two trainees and two other researchers. I was involved mostly in the first stages of study of the problem and the understanding of the codification of the model, as well as training the trainees on how to use the software and checking their results using different calibrations.

Software: AMPL.

Techniques: MTZ (Miller-Tucker-Zemlin formulation), Multivehicle Inventory Routing Problem (MIRP)

```
ampl MIRPfinal.mod
CPLEX 12.6.3.0: optimal integer solution within mipgap or absmipgap; objective 5815.72
18833126 MIP simplex iterations
647391 branch-and-bound nodes
absmipgap = 0.580225, relmipgap = 9.97683e-05
No basis.
fo = 5032.05

X [*,*,1,1]
:    0   1   2   3   4   5   6   7   8   9  10    :=
0    0   0   0   0   1   0   0   0   0   0   0
1    0   0   0   0   0   0   0   0   0   0   0
2    0   0   0   0   0   0   0   0   0   0   0
3    0   0   0   0   0   0   0   0   0   0   0
4    1   0   0   0   0   0   0   0   0   0   0
5    0   0   0   0   0   0   0   0   0   0   0
6    0   0   0   0   0   0   0   0   0   0   0
7    0   0   0   0   0   0   0   0   0   0   0
8    0   0   0   0   0   0   0   0   0   0   0
9    0   0   0   0   0   0   0   0   0   0   0
10   0   0   0   0   0   0   0   0   0   0   0

 [*,*,1,2]
:    0   1   2   3   4   5   6   7   8   9  10    :=
0    0   0   0   0   1   0   0   0   0   0   0
1    0   0   0   0   0   0   0   0   0   0   0
2    0   0   0   0   0   0   0   0   0   0   0
3    0   0   0   0   0   0   0   0   0   0   0
4    0   0   0   0   0   1   0   0   0   0   0
5    0   0   0   0   0   0   0   0   1   0   0
6    0   0   0   0   0   0   0   1   0   0   0
7    0   0   0   0   0   0   0   0   0   0   1
8    0   0   0   0   0   0   0   1   0   0   0   0
9    0   0   0   0   0   0   0   0   0   0   0   0
10   1   0   0   0   0   0   0   0   0   0   0   0
```

Tabla 32: Resultados política ML (H=3)

	Política ML (H=3)								
	General			MTZ			Flujos		
n	F.O.	Tiempo	GAP (%)	F.O.	Tiempo	GAP (%)	F.O.	Tiempo	GAP (%)
5	2108.34	0	0,00%	2108.34	0	0,00%	2108.34	0	0,00%
10	4510.61	0	0,00%	4510.61	1	0,00%	4510.61	0	0,00%
15	5589.7	12	0,00%	5589.7	5	0,00%	5589.7	10	0,00%
20	6859.02	3	0,00%	6859.02	2	0,00%	6859.02	1	0,00%
25	8227.86	420	0,00%	8227.86	113	0,00%	8227.86	43	0,00%
30	12088.9	358	1,00%	12066.9	556	0,00%	12066.90	30	0,00%
35	11790.5	1432	2,01%	11659.9	308	0,00%	11659.9	95	0,00%
40	13991.9	89	4,98%	13574.9	63	2,19%	13364.90	142	0,00%
45	14264.5	1532	1,98%	14179.1	490	0,00%	14179.1	125	0,00%
50	14577.3	198	2,48%	14577.3	3600	0,00%	14577.3	394	0,00%

5) Multiobjective optimization for solving the Multi-Depot Vehicle Routing Problem considering clusterization techniques:

Researcher (2015 - 2016): I did the full implementation.

Software: C++, MATLAB.

Techniques: ILO CPLEX, MDVRP heuristics, vehicles capacity and route duration infeasibility, multiobjective programming, ILS, matheuristics (exact model and heuristics).

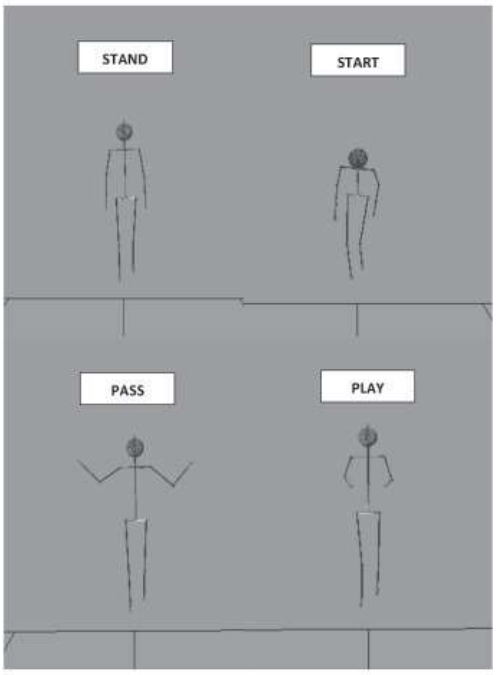
Tabla 2. Resultados instancias tipo p.

	Propuesta			**Vidal**			
caso	Promedio de 10 soluciones	Mejor tiempo requerido	Mejor solución	Mejor tiempo	Mejor solución - ILS	BKS	Gap
p01	576,87	0,22	576,87	0,49	576,87	576,87	0,00%
p02	473,53	0,23	473,53	0,81	473,53	473,53	0,00%
p03	641,18	0,62	641,18	0,99	640,65	640,65	0,08%
p04	1004,61	1,76	1001,03	0,81	999,21	999,21	0,18%
p05	750,93	2,68	750,02	2,09	750,03	750,03	0,00%
p06	882,76	1,88	878,10	1,6	876,5	876,5	0,18%
p07	887,45	3,56	884,66	1,63	881,97	881,97	0,31%
p08	4525,87	20,37	4500,70	14,85	4382,91	4375,49	2,86%
p09	3940,86	21,37	3933,99	12,75	3878,25	3859,17	1,94%
p10	3735,50	19,00	3713,54	14,43	3635,52	3631,11	2,27%
p11	3634,39	10,14	3624,66	12,94	3557,57	3546,06	2,22%
p12	1318,95	0,01	1318,95	1,09	1318,95	1318,95	0,00%
p13	1318,95	0,00	1318,95	0,98	1318,95	1318,95	0,00%
p14	1365,69	0,00	1365,69	0,98	1360,12	1360,12	0,41%
p15	2505,42	2,46	2505,42	3,25	2505,42	2505,42	0,00%
p16	2597,29	6,17	2597,29	3,15	2572,33	2572,23	0,97%
p17	2722,77	0,01	2722,77	3,03	2709,09	2709,09	0,50%
p18	3710,69	11,11	3707,77	10,32	3702,85	3702,85	0,13%
Promedio		5,66	2028,62	4,78	2007,82	2005,45	

	Propuesta			**Vidal**			
Instancia	Promedio 10 Soluciones	Mejor tiempo requerido	Mejor solución	Mejor tiempo	Mejor solución - ILS	BKS	Gap
pr01	861,32	0,21	861,32	1,02	861,32	861,32	0,00%
pr02	1296,25	1,55	1296,25	2,82	1296,25	1296,25	0,00%
pr03	1803,80	5,28	1803,8	5,66	1803,8	1803,8	0,00%
pr04	2076,72	8,22	2069,76	8,09	2045,45	2042,45	1,34%
pr05	2357,16	18,00	2350,69	11,41	2326,35	2324,45	1,13%
pr06	2715,26	28,56	2704,5	18,06	2668,76	2663,56	1,54%
pr07	1077,33	0,27	1077,33	1,55	1075,12	1075,12	0,21%
pr08	1658,23	5,56	1658,23	4,3	1658,23	**1658,23**	**0,00%**
pr09	2167,93	8,67	2154,82	8,85	2131,7	**2131,7**	1,08%
pr10	2911,79	34,23	2911,79	17,96	2810,25	**2805,53**	3,79%
Promedio		11,05	1,888,85	7,972	1867,7	**1866,24**	

6) Gamification of a Psychological Test for Assesing Self-Control in Students Using a Kinect Sensor:

Coresearcher (2015): I designed part of the experiment and processed the data, also adapted the sensor to be used. Worked with a psychology professor from another university and the data was collected from students.
Software and hardware: PEBL, FAAST, Kinect sensor, MATLAB.
Techniques: Iowa Gambling Test, 16-PF, Gamification, Exergames adaptation, Chi-square test.



g. 3. Gestures recognized by the FAAST as valid on the IGT performar

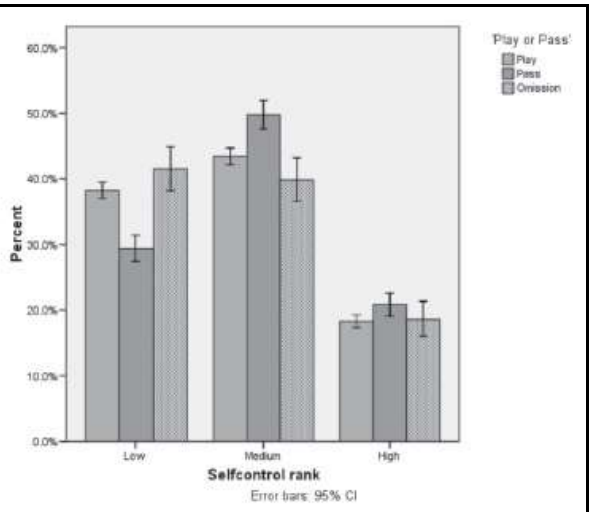
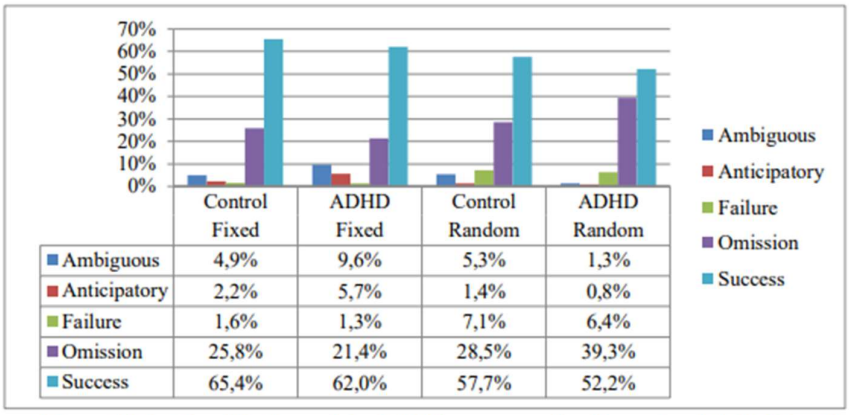


Fig. 5. Percentage of choices on the gamified IGT.

TABLE II
DESCRIPTIVE STATITCS

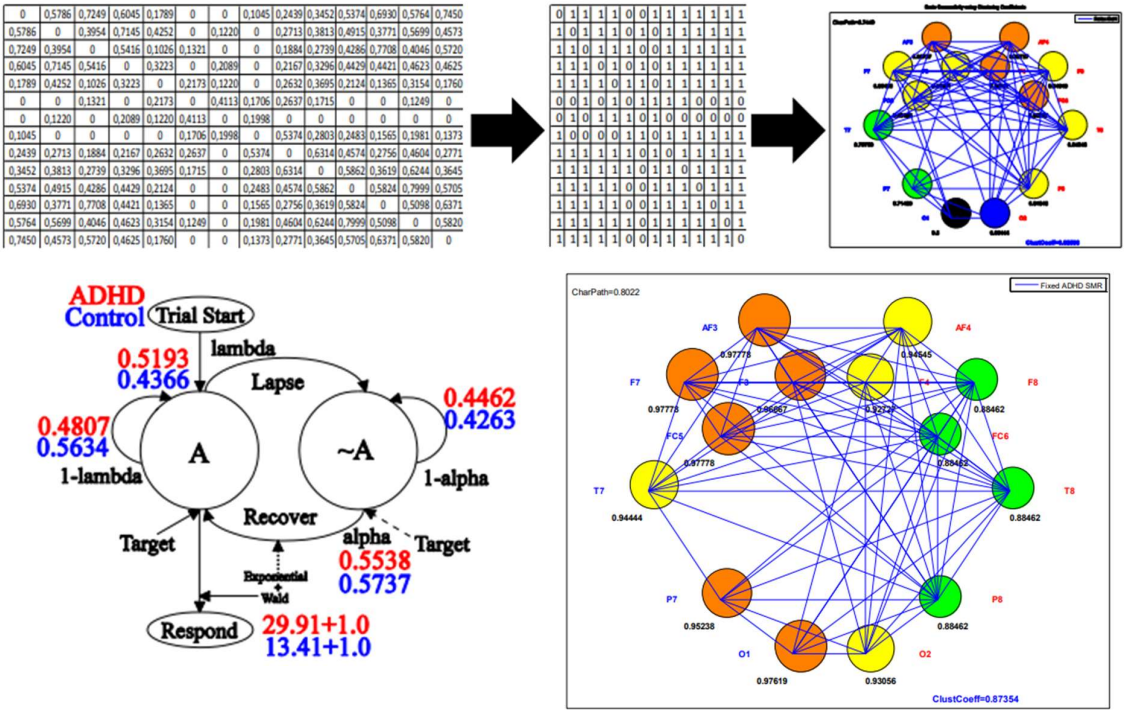
	Minimum	Maximum	Mean	Std. Deviation
Age	16	24	17.93	1.502
Self-control	1	10	4.34	2.232
'Reaction time'	.0001240	3.9991	1.2474	.7449
'Net change'	-1150	100	1.93	169.792

7) Design of an interactive system integrating virtual reality techniques and a BCI for evaluating and enhancing motor attention in ADHD patients: This was a project where I worked with people from sport sciences, physicians and physics engineers. I was involved in every step of the process and part of the results of this project were included in my masters thesis. We worked with Clinica del Dolor de Risaralda (Pain Clinic of Risaralda).
Young researcher funded by Colciencias (2014).
Software and Hardware: PEBL, R, MATLAB, Emotiv Epoc neuroheadset (EEG).
Techniques: Wavelet transform, Markov model, small-world network.



	Control Fixed	ADHD Fixed	Control Random	ADHD Random
Ambiguous	4,9%	9,6%	5,3%	1,3%
Anticipatory	2,2%	5,7%	1,4%	0,8%
Failure	1,6%	1,3%	7,1%	6,4%
Omission	25,8%	21,4%	28,5%	39,3%
Success	65,4%	62,0%	57,7%	52,2%

Figure 10.1. Classification of the SART responses for Control and ADHD participants.



2. CURRENT RESEARCH GROUP: GAOPE (Applications of Optimization Techniques and Stocastic Processes). I am coleading the research seedbed of GAOPE, our research group, which has the A1 score. That is the higher score given to a research group by the national administrative department of Science, Technology and Innovation Colciencias.

I have scientific papers published in journals and you can find those in my full CV or looking for my full name on Scholar Google.

3. CURRENT LECTURES.
I use the project-based learning approach in my classes.

1) Operations Research for Industrial Engineers: The subject focuses on integer and mixed linear programming models.
Content:
Linear programming standard problems.
Modeling linear problems using binary decision variables.
Branch and Bound algorithm.
Gomory cuts.
Dynamic Programming (including the inventory problem).
Multiobjective programming.
Heuristics and metaheuristics
Software: AMPL, QM, Concorde, LKH.

2) Data Analysis for Industrial Engineers. The subject focuses on giving students tools to be able to read external files, preprocess them, build charts and graphs, and being able to use those tools to storytelling.
Software: RStudio. **Packages:** tidyverse,rmarkdown, shiny.

3) Python Language for Electrical Engineers. The subject focuses on giving students tools to build implementations to solve engineering problems, that includes programming structures, exceptions handling, reading external files, plotting, object-oriented programming.
Software: PseInt, Spyder, Colab Google. **Packages:** numpy, pandas, matplotlib, scipy.