Ripply

WDL.

World Data League

Urban Data Lab Challenges

URBAN
CO-CREATION
DATA LAB

Agosto 2021

v1.0

Índice

Introdução **2**
- Desafios 2
- Contrapartidas da Ripply 3
- Responsabilidades do projeto UCD Lab 3

Descrição **4**
- Números 4
- Formato 4
- Agenda 5
- Prémios 6
- **Liga** **6**
  - Etapa 1 - Public Transportation 6
  - Etapa 2 - Traffic 7
  - Etapa 3 - Cycling 7
  - Etapa 4 - Environment 8
- **Finals - Noise** **10**

Challenges UCD Lab **11**

Conclusão **12**

ANEXOS **13**

Aviso Legal

Este documento é destinado exclusivamente para as pessoas a quem é dirigido, podendo conter informação confidencial. Qualquer reprodução total ou parcial, partilha ou uso impróprio deste conteúdo sem autorização prévia dos autores é expressamente proibido.

Introdução

Este documento serve como relatório relativamente à proposta de serviços de consultoria com vista à obtenção de insights e/ou soluções para determinados desafios reais do Projecto Urban Co-creation Data Lab[1] (UCD Lab), financiado pela Connecting Europe Facility (CEF) da EU e implementado pelo consórcio liderado pela UNL | NOVA IMS e que tem como parceiros a CML, a AMA, a NEC e a BSC-CNS (https://www.urbandatalab.pt/).

A World Data League (WDL) é uma competição que tem como base a exploração e análise de dados de modo a ajudar a resolver problemas de impacto social.
O objetivo desta iniciativa consiste em juntar as melhores mentes em ciência de dados, machine learning e inteligência artificial para construir soluções duradouras para esses desafios.

Desafios

Decidimos desafiar o projeto UCD Lab a fazer parte do evento WDL uma vez que há oportunidade de ajudar na possível resolução dos 5 seguintes desafios:

- **Mobilidade**[2] - Avaliação e descoberta de padrões e comportamentos da micro mobilidade na cidade de Lisboa.
- **Gestão de resíduos**[3] - Identificação de padrões / perfis e previsão da produção de resíduos sólidos na cidade de Lisboa.
- **Estacionamento**[4] - Identificação de padrões e predição do estacionamento ilegal na cidade de Lisboa permitindo otimizar a eficiência na fiscalização na cidade de Lisboa
- **Poluição**[5] - Elaboração de modelos preditivos para a propagação de poluidores líquidos e atmosféricos na cidade de Lisboa.
- **Emergência**[6] - desenvolver modelos preditivos para prever os acidentes de tráfego.

[1] https://urbandatalab.pt/index.php
[2] https://urbandatalab.pt/index.php/component/sppagebuilder/?view=page&id=116
[3] https://urbandatalab.pt/index.php/component/sppagebuilder/?view=page&id=117
[4] https://urbandatalab.pt/index.php/component/sppagebuilder/?view=page&id=118
[5] https://urbandatalab.pt/index.php/component/sppagebuilder/?view=page&id=119
[6] https://urbandatalab.pt/index.php/component/sppagebuilder/?view=page&id=122

Contrapartidas da Ripply

A seguinte lista de tópicos/tarefas contém os que a Ripply foi responsável por oferecer.

- Organizar e operacionalizar toda a competição;
- Garantir que pelo menos 2 desafios dos 5 acima mencionados são propostos durante a competição juntamente com os conjuntos de dados;
- Comunicar com antecedência e apoiar a equipa do UCD Lab de modo a fornecerem todo o material necessário para os desafios (descrição e dados);
- Disponibilizar as soluções dos participantes relativamente aos desafios (acima mencionados e utilizados no evento);
- Disponibilizar o white paper final do WDL,
- Enviar um relatório sumário do event.

Responsabilidades do projeto UCD Lab

Os seguintes tópicos foram da responsabilidade do projeto UCD Lab:

- Fornecer documentação detalhada dos desafios atempadamente;
- Fornecer os dados (datasets e documentação) atempadamente;
- Disponibilizar recursos humanos especializados para dar apoio técnico (análise de dados e/ou domínio) aos participantes sobre os desafios (mentores);
- Disponibilizar e/ou garantir uma equipa de suporte técnico para acesso aos dados durante o evento (pode ser de parceiros);
- Disponibilizar oradores/formadores para sessões de round-tables e/ou workshops.

Descrição

Números

Nota: Para consultar outras estatísticas, ver Anexos.

Inscrições

- 316 inscrições
- 44 nacionalidades

Participantes

- 186 participantes
- 34 nacionalidades

Equipas

- 50 equipas
- 13 países representados e 18 equipas internacionais[7]

Formato

A competição divide-se em duas grandes fases: Liga e Finais.

Liga

É constituída por 4 etapas, onde cada etapa está associada a um subtema do principal (*Data-driven cities)* e tem a duração de 2 semanas (Abril e Junho).

Em cada etapa são lançados desafios específicos. Cada equipa deverá escolher um desafio e submeter as soluções. No final desta fase, serão contabilizadas as 3 melhores submissões da equipa para determinar a pontuação final.

As 10 equipas com maior pontuação qualificam-se para a fase seguinte: as Finais.

[7] Uma equipa representa um país caso haja uma nacionalidade predominante (>50%).

Finais

Esta fase do evento divide-se, também, em duas subfases:

- 3 dias (60 horas) de competição (2 a 4 julho) - 10 equipas tentam resolver um dado desafio
- Apresentações em direto dos 3 finalistas (10 julho) - 3 melhores classificados pelo júri intermédio após a fase acima.

Agenda

Na tabela abaixo, constam as datas dos principais momentos do WDL.

15 de fevereiro a 17 de março de 2021	Fase de inscrições
16 a 31 de março de 2021	Seleção das equipas participantes
3 de abril de 2021	Início do WDL e Liga
3 a 17 de abril de 2021	Liga - Etapa 1 e workshop
17 de abril a 1 de maio de 2021	Liga - Etapa 2 e workshop
24 de abril de 2021	Anúncio vencedor Etapa 1 e Leaderboard
1 a 15 de maio de 2021	Liga - Etapa 3 e workshop
8 de maio de 2021	Anúncio vencedor Etapa 2 e Leaderboard
15 a 29 de maio 2021	Liga - Etapa 4 e workshop
22 de maio de 2021	Anúncio vencedor Etapa 4 e Leaderboard
5 de junho de 2021	Anúncio vencedor Etapa 4, 10 finalistas e Leaderboard geral da Liga
6 de junho a 1 de julho de 2021	Preparação dos finalistas
2 a 4 de julho de 2021	Finais
7 de julho de 2021	Anúncio do Top 3
9 de julho de 2021	Preparação para as apresentações finais
10 de julho de 2021	Apresentações ao vivo do Top 3 e anúncio do vencedor
26 de setembro de 2021	Prazo para publicar o white paper

Prémios

O valor total de prémios foi de 15.000€, distribuídos da seguinte forma:

- 1º Lugar: 7.000€
- 2º Lugar: 4.000€
- 3º Lugar: 1.500€
- Cada equipa finalista: 100€
- Vencedor de cada etapa: 250€
- Outras atividades: 500€

Liga

Etapa 1 - Public Transportation

- Link das soluções: aqui
- Equipas participantes: 50
- Soluções submetidas: 26

Position	Team	Country	Points
01 🥇	Tech Moguls	Portugal	209
02 🥈	Data Sloths	International	201
03 🥉	Children of Jupyter	Portugal	176

Desafio 1 - *Churn model for public transports*

- Fornecedor: PSE (Portugal)
- Nº submissões: 15

Desafio 2 - *Model of integrated transports for senior citizens*

- Fornecedor: PSE (Portugal)
- Nº submissões: 11

Etapa 2 - Traffic

- Link das soluções: aqui
- Equipas participantes: 47
- Soluções submetidas: 25

Position	Team	Country	Points
01 🥇	Data Sloths	International	238
02 🥈	Tech Moguls	Portugal	234
03 🥉	AKITA	China	218

Desafio 1 - *Predicting traffic flow in a city using induction loop sensors*

- Fornecedor: Municipality of Porto / Porto Digital (Portugal)
- Nº submissões: 12

Desafio 2 - *Identification of patterns and prediction of irregular parking*

- Fornecedor: **UCD Lab** (Portugal)
- Nº submissões: 3

Desafio 3 - *Patterns and predictive modelling of traffic accidents*

- Fornecedor: **UCD Lab** (Portugal)
- Nº submissões: 6

Desafio 4 - *Identifying road segments with potential safety hazard*

- Fornecedor: PSE (Portugal)
- Nº submissões: 4

Etapa 3 - Cycling

- Link das soluções: aqui
- Equipas participantes: 41
- Soluções submetidas: 15

Position	Team	Country	Points
01 🥇	Data Sloths	International	237
02 🥈	Old Walruses	Hungary	237
03 🥉	Children of Jupyter	Portugal	207

Desafio 1 - *Predicting the demand for shared bicycles*

- Fornecedor: Municipality of Valle de Aburrá (Colombia)
- Nº submissões: 9

Desafio 2 - *(Literally) Paving the way towards safer cities*

- Fornecedor: CycleAI (Portugal)
- Nº submissões: 4

Desafio 3 - *Missing Links - Closing the circuit in existing cycle networks*

- Fornecedor: U-Shift (Portugal)
- Nº submissões: 2

Etapa 4 - Environment

- Link das soluções: aqui
- Equipas participantes: 38
- Soluções submetidas: 11

Position	Team	Country	Points
01 🥇	Tech Moguls	Portugal	270
02 🥈	Data Sloths	International	243
03 🥉	Money House	Portugal	209

Desafio 1 - *Strategies to reduce the environmental and health impacts of pollution*

- Fornecedor: PSE (Portugal)
- Nº submissões: 0

Desafio 2 - *Attracting population to green spaces in metropolitan areas*

- Fornecedor: PSE (Portugal)
- Nº submissões: 5

Desafio 3 - *Optimisation of the outdoor advertising in cities*

- Fornecedor: PSE (Portugal)
- Nº submissões: 4

Desafio 4 - *Predicting air quality for outdoor activities*

- Fornecedor: Municipality of Cascais (Portugal)
- Nº submissões: 2

Leaderboard - Liga

Position	Team	Country	Points
01 🥇	Data Sloths	International	718
02 🥇	Tech Moguls	Portugal	713
03 🥇	Children of Jupyter	Portugal	596
04 🥇	NextEpoch	International	575
05 🥇	AKITA	China	564
06 🥇	i bellissimi di rete 4	Italy	567
07 🥇	Sidereus	International	550
08 🥇	DataChics	Spain	547
09 🥇	urbanata	India	544
10 🥇	BelManunel Soto	International	541

Finals - Noise

- Link das soluções: *aqui*
- Equipas participantes: 10
- Soluções submetidas: 10

Desafio - *Improving the quality of life by reducing city noise levels*

- Fornecedor: Città Di Torino (Itália)

Leaderboard Final

1. Data Sloths	245 pts
2. Tech Moguls	242 pts
3. Children of Jupyter	236 pts
4. NextEpoch	214 pts
5. AKITA	200 pts
6. BelManunel Soto	168 pts
7. DataChics	167 pts
8. i bellissimi di rete 4	142 pts
9. Sidereus	130 pts
10. urbanata	119 pts

Challenges UCD Lab

Como destacado anteriormente, durante a WDL foram disponibilizados dois desafios do UCD Lab durante a segunda etapa (*Traffic*) da fase denominada Liga.

Os desafios consistiram nos seguintes:

- **Desafio 2 - Identification of patterns and prediction of irregular parking** *(aqui)*

  Três equipas submeteram possíveis soluções: Children of Jupyter (5º), Tekorner (11º) e Team DS Alpha (12º).

- **Desafio 3 - Patterns and predictive modelling of traffic accidents** *(aqui)*

  Seis equipas submeteram possíveis soluções: AKITA (3º), Money House (6º), i belissimi di rete 4 (8º), Sustainable DTrees (14º), Old Walruses (20º) e DSinGoodHands (21º),

Leaderboard

Position	Team	Country	Points
01 🥇	Data Sloths	International	238
02 🥈	Tech Moguls	Portugal	234
03 🥉	AKITA	China	218
04	DataChics	Spain	216
05	Children of Jupyter	Portugal	213
06	Money House	Portugal	196
07	BelManunel Soto	International	192
08	i bellissimi di rete 4	Italy	190
09	urbanata	India	181
10	NextEpoch	India	176

Todos os insights retirados destas soluções estão presentes no *white paper* que será disponibilizado ao UCD Lab, como indicado na proposta.

Conclusão

O WDL pretende ser uma competição que junta as melhores mentes em ciência de dados, inteligência artificial e *machine learning* de modo a resolver problemas de impacto social.

Esta primeira edição consistiu em dar início a um projeto que se pretende que seja de longo-prazo. Foi necessário definir todo o formato, regras e outros procedimentos. Assim como foi também necessário fazer o levantamento do ecossistema relacionado com o tópico principal desta edição - Data-driven cities.

Tendo em consideração o exposto acima, foi possível disponibilizar diversos desafios organizados por subtemas relacionados com o 11º SDG[8]. A parceria efetuada com as várias entidades especialistas nos temas trabalhados revelou-se fundamental para a boa definição dos desafios e o devido suporte que foi necessário disponibilizar aos participantes.

Em relação às soluções apresentadas, como seria de prever, o número das mesmas foi diminuindo com o avançar da Liga. Ainda assim, apenas um dos desafios não teve qualquer submissão. Estas soluções encontram-se todas disponíveis publicamente e podem ser utilizadas segundo licença MIT[9].

No que diz respeito aos desafios propostos pelo UCD Lab, foi possível integrá-los no contexto da competição e adquirir alguns insights sobre os mesmos, descritos no white paper.

Apesar do sucesso nesta primeira edição, que serviu como primeiro teste a toda a estrutura e formato, considera-se que ainda existem muitos pontos por melhorar em futuras edições, mas que estão construídas bases sólidas para fazer da WDL um evento à escala mundial e com impacto real na sociedade.

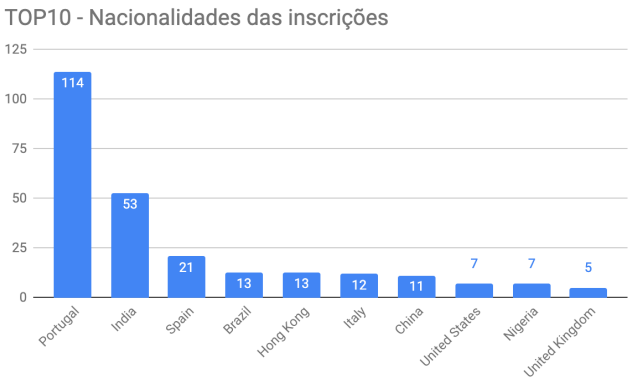
Uma das ideias em futuras edições é propor os mesmos desafios (ou ligeiramente melhorados) para que as equipas tenham como base as soluções já propostas e consigam, deste modo, chegar a melhores resultados e obter ainda mais insights.

Finalmente, um muito obrigado ao UCD Lab, e a todas as pessoas que o representaram e que ajudaram a tornar este evento um sucesso.

[8] https://www.globalgoals.org/11-sustainable-cities-and-communities
[9] https://pt.wikipedia.org/wiki/Licen%C3%A7a_MIT

ANEXOS

Anexo 1 - Inscrições



GENDER STATS PER APPLICANT		
Female	86	27%
Male	228	73%

AGE STATS PER APPLICANT		
18-25	191	61%
26-35	102	33%
36-45	18	6%
46 - 60	0	0%
>60	0	0%

FIELD OF WORK STATS PER APPLICANT		
Data Analytics / Business Intelligence	70	22%
Team Management	4	1%
Quality Assurance / Testing	2	1%
Student	93	29%
Machine Learning / Deep Learning	89	28%
Data Engineering	5	2%
Software Development / Programming	21	7%
Teaching	2	1%
Product Management	3	1%
Research	18	6%
Other	11	3%

Anexo 2 - Participantes

GENDER STATS PER APPLICANT		
Female	61	33%
Male	125	67%

AGE STATS PER APPLICANT		
18-25	92	49%
26-35	77	41%
36-45	15	8%
46-60	2	1%
>60	0	0%

FIELD OF WORK STATS PER APPLICANT	
Data Analytics / Business Intelligence	42
Data Engineering	3
Machine Learning / Deep Learning	60
Other	6
Product Management	2
Quality Assurance / Testing	1
Research	16
Software Development / Programming	7
Student	46
Teaching	2
Team Management	1

YEAR EXPERIENCE PER APPLICANT		
None	16	9%
<1	35	19%
1-3	70	38%
4-6	36	19%
7-15	19	10%
>15	10	5%

FIELD OF STUDY PER APPLICANT	
Biology / Biochemistry	5
Data Analytics / Business Intelligence	36
Economy / Finance	7
Machine Learning / Deep Learning	55
Marketing	2
Mathematics / Statistics	12
Other	3
Other Engineering	26
Physics	5
Software Engineering / Computer Science	35

YEAR DS/AI/ML STATS		
None	7	4%
<1	52	28%
1-3	100	54%
4-6	21	11%
7-15	5	3%
>15'	1	1%

NR HACKATONS PER APPLICANT		
None	86	46%
1-2	68	37%
3-5	25	13%
>5	7	4%

Anexo 3 - Equipas

TEAMS STATS	
Argentina	1
Belgium	1
Brazil	1
China	2
Colombia	1
Hong Kong	1
Hungary	1
India	7
International	18
Italy	2
Nigeria	2
Portugal	8
Spain	4
UK	1
TOTAL	**50**

TEAMS THAT FORFEITED		
Pre-Stage	0	0%
Stage 1	3	6%
Stage 2	6	12%
Stage 3	3	6%
Stage 4	0	0%
TOTAL	**12**	**24%**

NATIONALITIES STATS PER APPLICANT		
Portugal	46	12%
India	30	8%
Spain	20	5%
Italy	12	3%
China	9	2%
Brazil	8	2%
Nigeria	6	2%
Hong Kong	5	1%
United Kingdom	5	1%
United States	5	1%
Argentina	4	1%
Hungary	4	1%
Belgium	3	1%
Colombia	3	1%
Germany	3	1%
Albania	2	1%
France	2	1%
Greece	2	1%
Morocco	2	1%
Algeria	1	0%
Australia	1	0%
Austria	1	0%
Egypt	1	0%
Indonesia	1	0%
Ireland	1	0%
Kazakhstan	1	0%
Nepal	1	0%
Norway	1	0%
Poland	1	0%
Romania	1	0%
Russia	1	0%
South Africa	1	0%
Tunisia	1	0%
Turkey	1	0%
TOTAL	**186**	