eleven | ENPC – Département IMI Hackaton – The Endless Line

To the attention of IMI Department

February 13th, 2024

The Endless Line

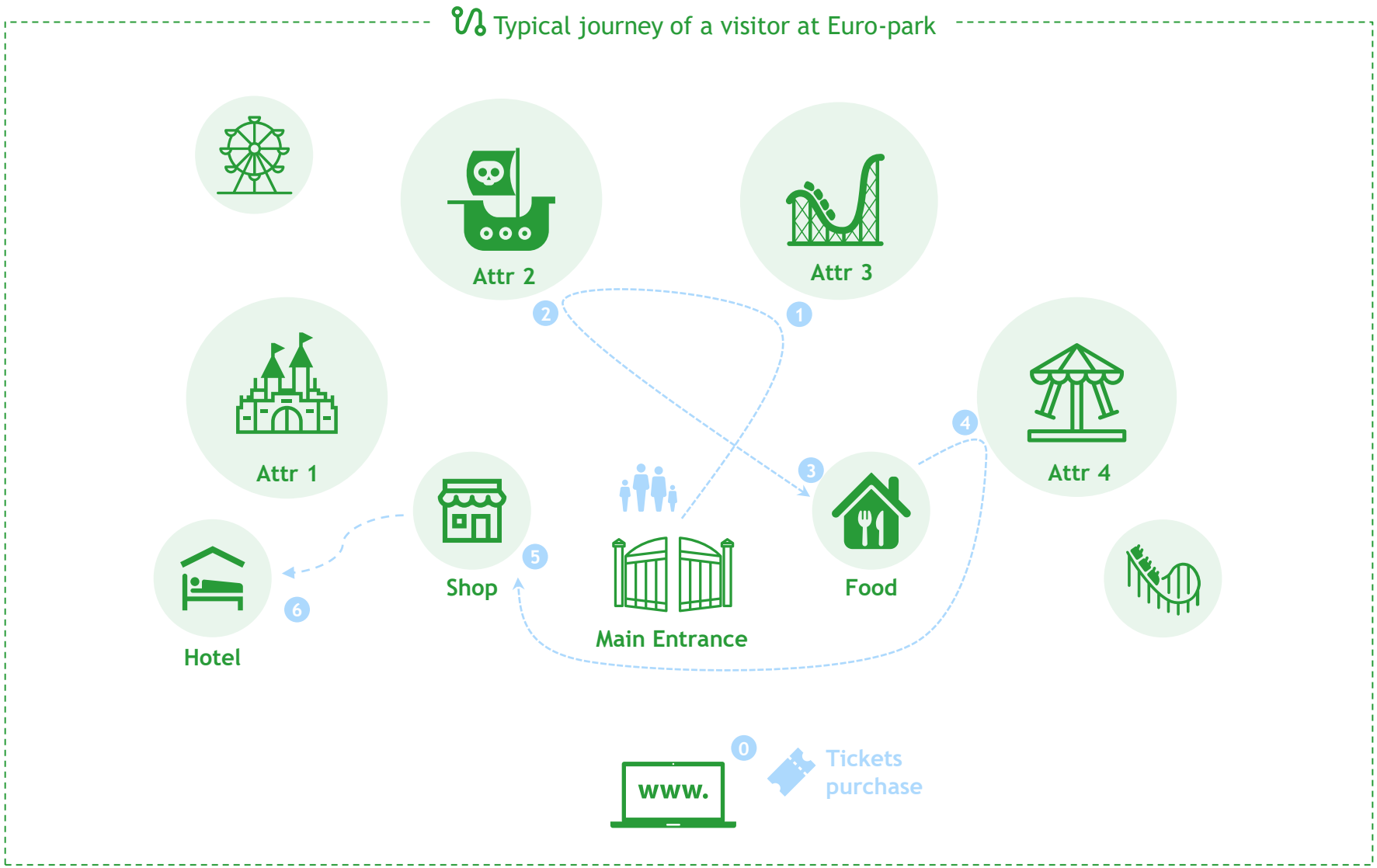
Forecasting waiting times in a theme park to improve visitor experience

eleven's challenge: find the best way to model amusement parks lines leveraging your expertise in data science

Exercise:

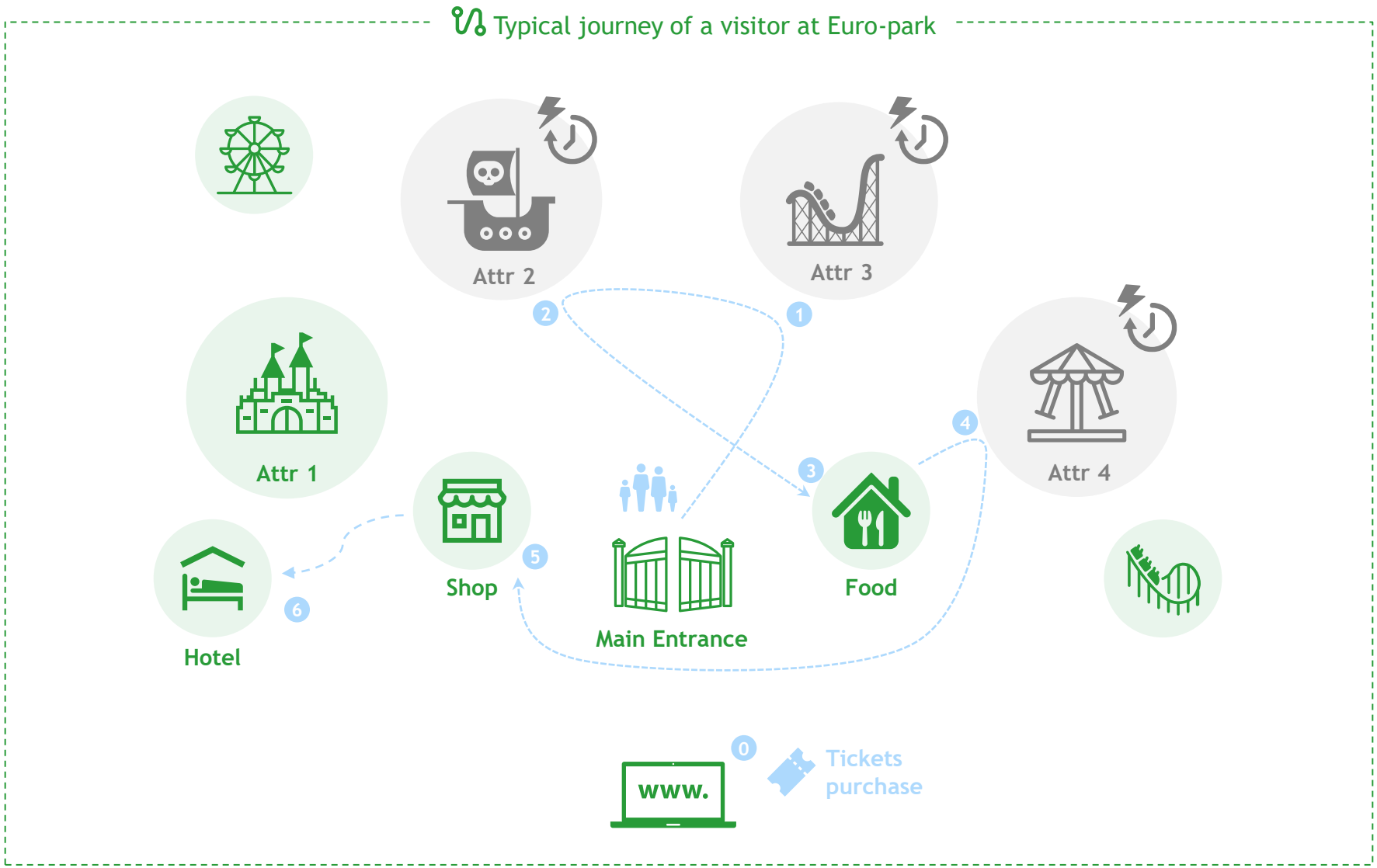
- ✓ Make groups of 4 people
- ✓ On your topic, put yourself in a data consultant's shoes during the exploration of a new use-case:
  - ✓ What is the structure of the data I can access? What are its main trends ?
  - ✓ What is the best way to model my data and the phenomenon I want to predict ?
- ✓ Develop a model and a pipeline to predict waiting times on the amusement park's main rides
- ✓ Test your model on the validation data using the provided leaderboard
- ✓ Compare to other groups on a hidden test set
- ✓ Present to the jury and prepare for questions

Context: Euro-park, a global theme park, is experiencing a significant increase of waiting times at attractions, which affects visitor satisfaction (1/3)



Sources: eleven strategy

Context: Euro-park, a global theme park, is experiencing a significant increase of waiting times at attractions, which affects visitor satisfaction (2/3)

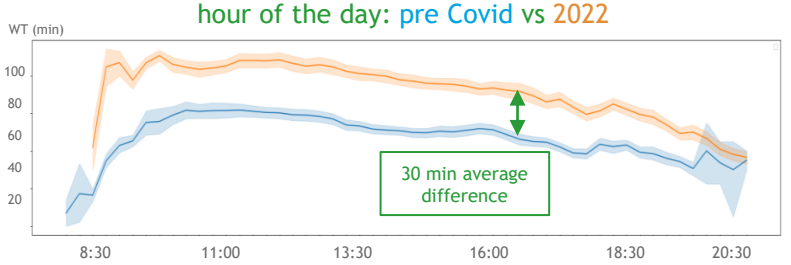


Sources: eleven strategy

Context: Euro-park, a global theme park, is experiencing a significant increase of waiting times at attractions, which affects visitor satisfaction (3/3)

Typical journey of a visitor at Euro-park

Average waiting time at attraction 3 at every hour of the day: pre Covid vs 2022



Major Pain point: significant increase of waiting times since pre-covid era resulting in degraded visitor experience and dissatisfaction

Attr 3

Attr 4

Food

Shop

Main Entrance

Hotel

www.

Tickets purchase

Sources: eleven strategy

Objectives: Euro-park, therefore, wants to accurately forecast waiting times for its attractions and identify use cases to leverage this information to improve the park's KPIs

Scope of the case study

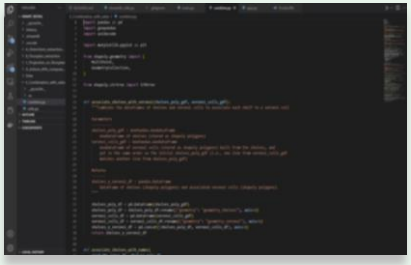
1

How to accurately forecast attractions waiting times?

2

How to leverage this information to enhance Euro-park's KPIs?

Deliverables



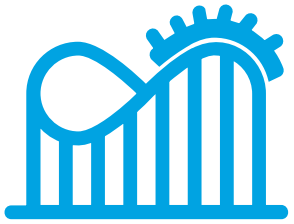
1. The code used to solve the case and a short presentation

2. A test on a hidden test set to compare with other teams

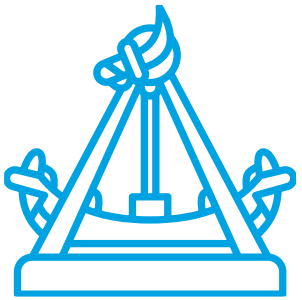
Sources: eleven strategy

Objectives: predict the waiting times of three main attractions 2 hours in advance

Euro-park operates hundreds of attractions in the park: in this case study, we will focus on waiting times of three of them



Flying Coaster



Pirate Ship

Water Ride

4 files are available for the analysis:

waiting_times_train.csv

1

waiting_times_X_test_val.csv

2

waiting_times_X_test_final.csv

3

weather.csv

4

Data available: past waiting times are provided for each attraction at a granularity of 15 minutes. In the train set, you have access to the prediction variable : WAIT_TIME_IN_2H

The prediction variable

	A	B	C	D	E	F	G	H	I
1	DATETIME	ENTITY_DESCRIPTION_SHORT	ADJUST_CAPACITY	DOWNTIME	CURRENT_WAIT_TIME	TIME_TO_PARADE_1	TIME_TO_PARADE_2	TIME_TO_NIGHT_SHOW	WAIT_TIME_IN_2H
2	05/02/2022 11:45	Water Ride	247.0	0	20				30.0
3	24/02/2019 10:45	Water Ride	247.0	0	30	375.0		495.0	25.0
4	17/07/2021 15:45	Pirate Ship	280.5	0	35				35.0
5	03/04/2022 19:45	Pirate Ship	230.35	0	15	-135.0		195.0	10.0
6	20/10/2021 10:30	Pirate Ship	153.0	0	15				10.0
7	30/05/2022 09:15	Flying Coaster	176.4	0	5	495.0		825.0	10.0
8	22/09/2019 11:45	Water Ride	247.0	0	15	345.0		555.0	30.0
9	10/01/2019 14:30	Water Ride	247.0	0	15	120.0		210.0	15.0
10	20/08/2020 10:30	Water Ride	247.0	0	20				25.0
11	03/02/2022 16:00	Pirate Ship	148.75	0	30				20.0
12	28/08/2019 19:30	Flying Coaster	756.0	0	25	-120.0		210.0	15.0
13	30/09/2020 12:00	Pirate Ship	119.0	0	10				20.0
14	19/02/2019 10:45	Water Ride	224.5	0	15	375.0		495.0	25.0
15	15/12/2018 09:45	Flying Coaster	756.0	0	10	435.0	125.0	735.0	40.0
16	30/10/2021 13:00	Pirate Ship	306.0	0	15				35.0
17	03/04/2022 09:45	Pirate Ship	153.0	0	15	465.0		795.0	45.0
18	15/06/2021 09:30	Pirate Ship	52.7	0	5				15.0
19	06/07/2022 14:45	Water Ride	247.0	0	25	165.0		495.0	25.0
20	07/09/2019 12:15	Flying Coaster	756.0	0	40	315.0		585.0	50.0
21	13/10/2018 11:00	Water Ride	224.5	0	30	390.0	70.0	660.0	25.0
22	23/10/2020 17:00	Water Ride	247.0	0	20				10.0
23	14/12/2019 12:00	Flying Coaster	756.0	0	30	300.0	0.0	600.0	45.0
24	27/02/2020 13:30	Flying Coaster	756.0	0	25	210.0		390.0	35.0
25	01/05/2019 11:30	Flying Coaster	756.0	0	40	360.0		630.0	45.0
26	28/08/2019 17:00	Water Ride	247.0	0	35	30.0		360.0	25.0
27	02/10/2020 14:15	Flying Coaster	756.0	0	5				5.0
28	27/02/2019 11:00	Water Ride	247.0	0	35	360.0		510.0	20.0

Datetime: studied time
Entity_description_short: the name of the concerned attraction
Adjust_capacity: the real time capacity of the attraction (ie number of passengers)
Downtime: time in minutes when the attraction is down during the 15 next minutes
Current_wait_time: the current waiting times
Time_to_parade_1, 2, night_show: the time in minutes before the event of the concerned day
Wait_time_in_2H: the variable we want to predict

Data available: past waiting times are provided for each attraction at a granularity of 15 minutes. In the validation and final set, you only have access to the features

	A	B	C	D	E	F	G	H
1	DATETIME	ENTITY_DESCRIPTION_SHORT	ADJUST_CAPACITY	DOWNTIME	CURRENT_WAIT_TIME	TIME_TO_PARADE_1	TIME_TO_PARADE_2	TIME_TO_NIGHT_SHOW
2	23/11/2019 10:45	Water Ride	247.0	0	20	375.0	75.0	675.0
3	03/01/2022 16:45	Pirate Ship	153.0	0	45			
4	04/12/2021 15:30	Pirate Ship	255.0	0	40			
5	05/02/2020 13:15	Water Ride	247.0	0	15	225.0		345.0
6	13/05/2022 15:15	Flying Coaster	756.0	0	35	135.0		465.0
7	04/01/2020 10:00	Pirate Ship	221.0	0	50	420.0	120.0	720.0
8	03/10/2020 14:00	Water Ride	247.0	0	15			
9	26/06/2021 10:45	Flying Coaster	756.0	0	20			
10	11/11/2018 19:30	Flying Coaster	756.0	0	5	-150.0	-460.0	90.0
11	10/10/2018 14:15	Flying Coaster	756.0	0	25	195.0	-125.0	345.0
12	22/04/2019 19:15	Water Ride	247.0	0	20	-105.0		165.0
13	03/07/2019 11:45	Water Ride	247.0	0	20	345.0		675.0
14	07/10/2020 13:45	Flying Coaster	756.0	0	5			
15	17/07/2022 12:30	Water Ride	247.0	0	25	300.0		630.0
16	05/05/2022 16:45	Pirate Ship	306.0	0	35	45.0		375.0
17	08/03/2020 12:45	Water Ride	247.0	0	25	255.0		435.0
18	06/10/2018 16:30	Pirate Ship	280.5	0	70	60.0	-260.0	330.0
19	05/06/2022 15:00	Flying Coaster	756.0	0	50	150.0		480.0
20	29/09/2019 17:15	Flying Coaster	756.0	0	5	15.0	-285.0	225.0
21	06/09/2019 18:30	Flying Coaster	756.0	0	25	-60.0		150.0
22	20/06/2021 09:45	Water Ride	247.0	0	5			
23	01/03/2022 10:15	Pirate Ship	303.733	0	30			

One file for the validation step:

You can try as much as you want

One file for the final evaluation step:

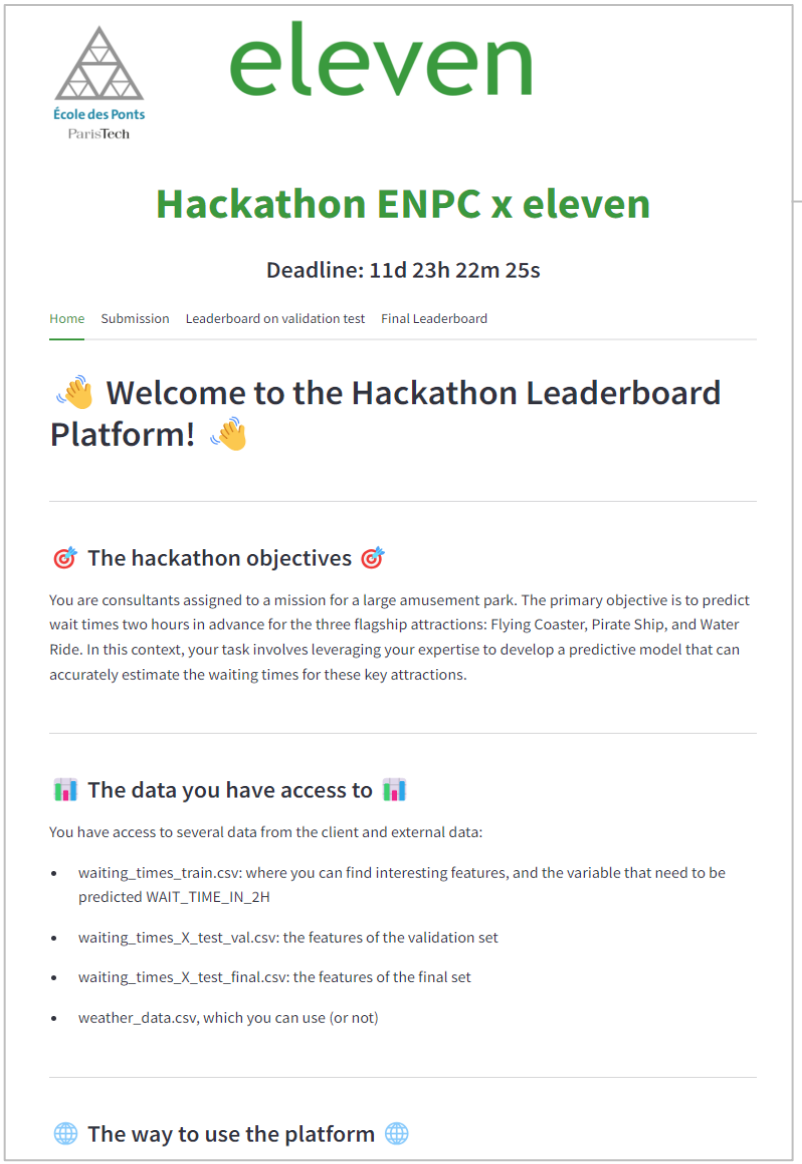
One final prediction

Data available: weather information are provided, for the past and future. It is up to you to use them or not

	A	B	C	D	E	F	G	H	I	J
1	temp	dew_point	feels_like	pressure	humidity	wind_speed	rain_1h	snow_1h	clouds_all	DATETIME
2	12.17	7.68	11.37	1019.0	74.0	3.3			100.0	01/10/2018 00:00
3	11.995000000000001	7.6075	11.192499999999999	1019.0	74.5	3.2624999999999997			99.25	01/10/2018 00:15
4	11.82	7.535	11.015	1019.0	75.0	3.2249999999999996			98.5	01/10/2018 00:30
5	11.645	7.4624999999999995	10.8375	1019.0	75.5	3.1875			97.75	01/10/2018 00:45
6	11.47	7.39	10.66	1019.0	76.0	3.15			97.0	01/10/2018 01:00
7	11.445	7.4624999999999995	10.645	1019.0	76.5	3.12			97.25	01/10/2018 01:15
8	11.42	7.535	10.629999999999999	1019.0	77.0	3.09			97.5	01/10/2018 01:30
9	11.395	7.6075	10.615	1019.0	77.5	3.0599999999999996			97.75	01/10/2018 01:45
10	11.37	7.68	10.6	1019.0	78.0	3.03			98.0	01/10/2018 02:00
11	11.23	7.635	10.459999999999999	1019.0	78.5	3.1025			96.5	01/10/2018 02:15
12	11.09	7.59	10.32	1019.0	79.0	3.175			95.0	01/10/2018 02:30
13	10.95	7.545	10.18	1019.0	79.5	3.2474999999999996			93.5	01/10/2018 02:45
14	10.81	7.5	10.04	1019.0	80.0	3.32			92.0	01/10/2018 03:00
15	10.9	7.5875	10.1375	1019.0	80.0	3.4025			88.25	01/10/2018 03:15
16	10.99	7.675	10.235	1019.0	80.0	3.485			84.5	01/10/2018 03:30
17	11.08	7.762499999999999	10.3325	1019.0	80.0	3.5675			80.75	01/10/2018 03:45
18	11.17	7.85	10.43	1019.0	80.0	3.65	0.15		77.0	01/10/2018 04:00
19	11.219999999999999	7.99	10.497499999999999	1019.0	80.5	3.8425	0.175		73.5	01/10/2018 04:15
20	11.27	8.129999999999999	10.565	1019.0	81.0	4.035	0.2		70.0	01/10/2018 04:30
21	11.32	8.27	10.6325	1019.0	81.5	4.2275	0.225		66.5	01/10/2018 04:45
22	11.37	8.41	10.7	1019.0	82.0	4.42	0.25		63.0	01/10/2018 05:00
23	11.295	8.3375	10.6175	1019.25	82.0	4.5375	0.24416666666666667		57.0	01/10/2018 05:15
24	11.219999999999999	8.265	10.535	1019.5	82.0	4.654999999999999	0.23833333333333334		51.0	01/10/2018 05:30
25	11.145	8.192499999999999	10.452499999999999	1019.75	82.0	4.7725	0.23249999999999998		45.0	01/10/2018 05:45
26	11.07	8.12	10.37	1020.0	82.0	4.89	0.22666666666666666		39.0	01/10/2018 06:00
27	10.785	7.704999999999999	9.657499999999999	1020.0	81.25	4.9025	0.22083333333333333		35.25	01/10/2018 06:15
28	10.5	7.289999999999999	8.945	1020.0	80.5	4.915	0.215		31.5	01/10/2018 06:30

You have access to weather information in the localisation of the park, actualized every 15 minutes

To assess you model's performance and compare results between groups, a leaderboard is available for the duration of the hackathon



École des Ponts ParisTech

eleven

Hackathon ENPC x eleven

Deadline: 11d 23h 22m 25s

Home | Submission | Leaderboard on validation test | Final Leaderboard

👋 Welcome to the Hackathon Leaderboard Platform! 👋

🎯 The hackathon objectives 🎯

You are consultants assigned to a mission for a large amusement park. The primary objective is to predict wait times two hours in advance for the three flagship attractions: Flying Coaster, Pirate Ship, and Water Ride. In this context, your task involves leveraging your expertise to develop a predictive model that can accurately estimate the waiting times for these key attractions.

📊 The data you have access to 📊

You have access to several data from the client and external data:

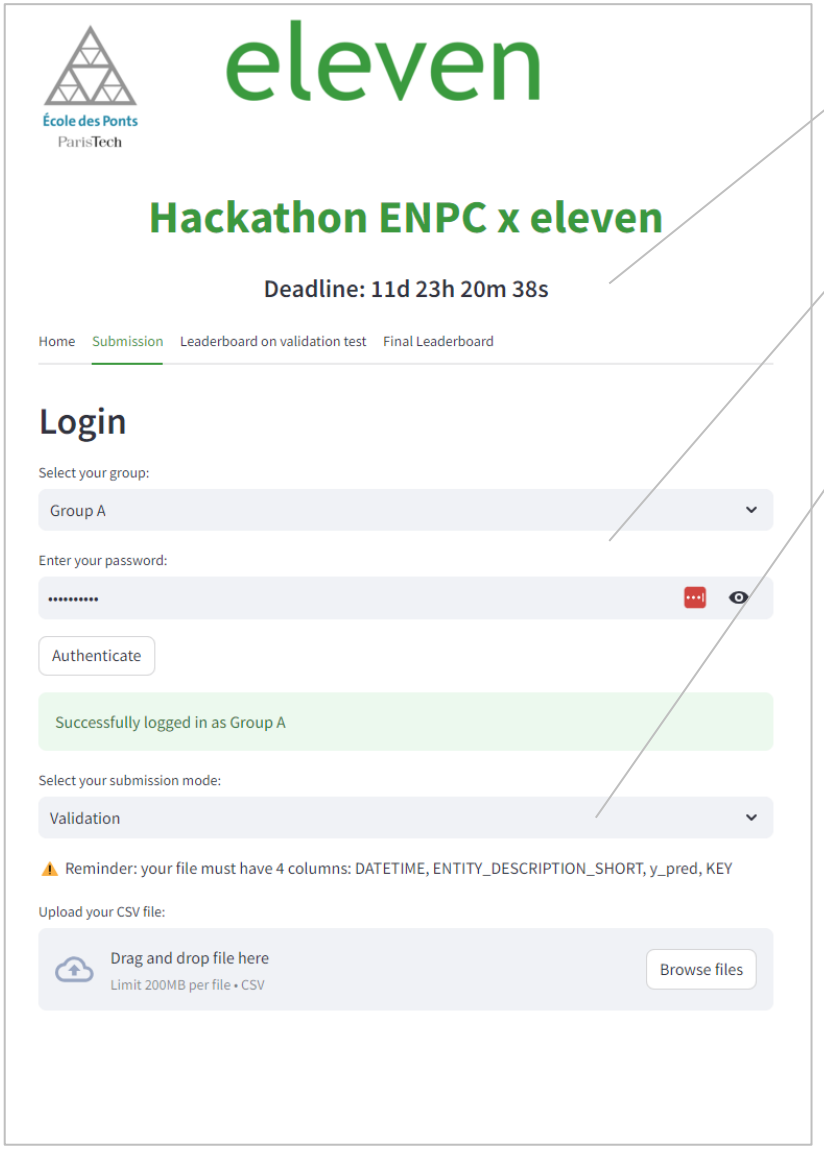
- waiting_times_train.csv: where you can find interesting features, and the variable that need to be predicted WAIT_TIME_IN_2H
- waiting_times_X_test_val.csv: the features of the validation set
- waiting_times_X_test_final.csv: the features of the final set
- weather_data.csv, which you can use (or not)

🌐 The way to use the platform 🌐

Welcome page, that restates the objectives & available data

[Streamlit (hackatonenpctheendlessline.streamlit.app)](hackatonenpctheendlessline.streamlit.app)

Each group will be provided a password to submit its answer during the hackathon, and for the final submission



Check the time left for the hackathon

Authenticate using the provided login and password

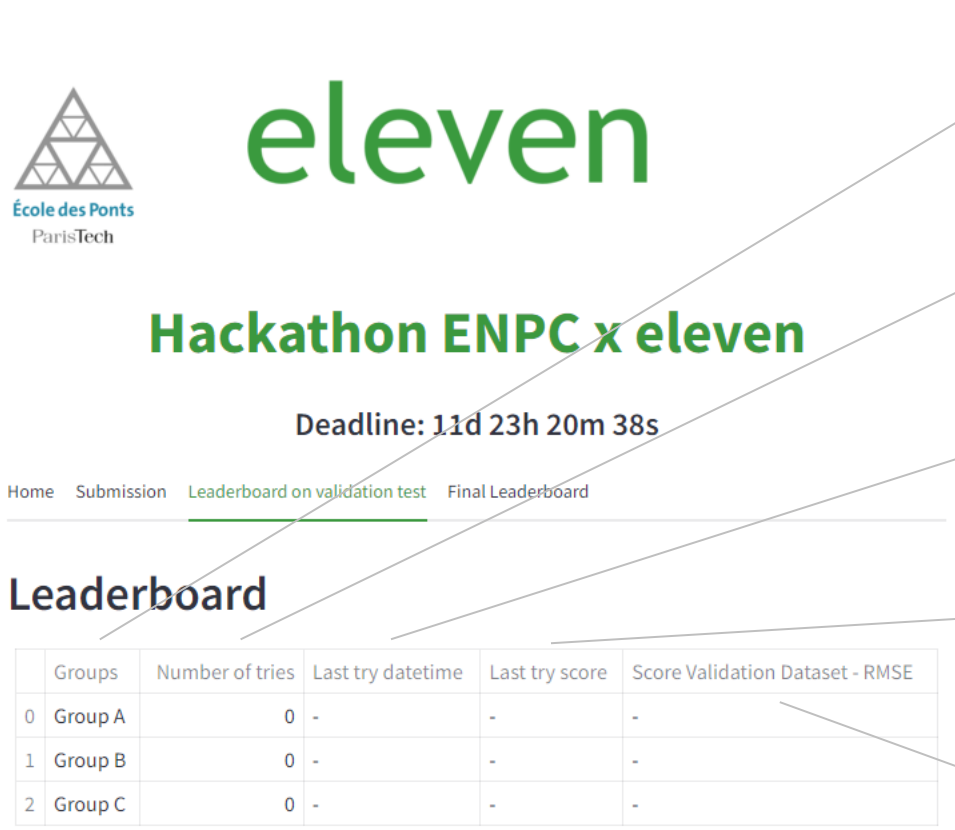
Select the type of submission (“Validation” to test your algorithms,
“Final” for final predictions on the hidden set)

	A	B	C	D	E
1	DATETIME,ENTITY_DESCRIPTION_SHORT,y_pred,KEY				
2	2019-11-23 10:45:00,Water Ride,150000.0,Validation				
3	2022-01-03 16:45:00,Pirate Ship,150000.0,Validation				
4	2021-12-04 15:30:00,Pirate Ship,150000.0,Validation				

Submissions are expected to be under commas separated .csv format, with four columns DATETIME, ENTITY_DESCRIPTION_SHORT and y_pred, and KEY as shown

KEY value for the final submission will be provided on Wednesday afternoon. Be careful, only one try is allowed!

Each group will be provided a password to submit its answer during the hackathon, and for the final submission



Groupe name

Number of tries that were made by your team

Datime of you last attempt

Score of you last attempt

Your best attempt so far

The chosen metric is the Root Mean Squared Error (RMSE) of your predictions

Proposed schedule for the week: the timing may be short, do not hesitate to split the work between the members of the team

	Tuesday 13th	Wednesday 14st
AM session	Presentation of eleven (*9:00 - 9:15*) Presentation of eleven's assignments *(9:15 - 10:15)*	Online Q&A (*9: 00 - 10: 00*)
	Refresher on Supervised Learning (*10:30 - 11:30*)	Working Time (*10: 00 - 12 : 00*)
	Hackaton Presentation and Kick-Off (*11:30 - 12:00*)	
PM session	Working Time (*13: 00 - 14 : 00*)	Working Time (*13: 00 - 16 : 00*)
	Q&A, Set-Up verifications (*14: 00 - 15 : 00*)	Jury (*16:00 - 17: 00*)
	Working Time (*15: 00 - 18 : 00*)	Closing Ceremony (*17: 00 - 17: 30*)

Final presentation details and best practices:

On Wednesday, you will have to present your work in front of a jury during a closed-door session

The modalities of the presentation will be as follow:

- 5min group pitch, showing the main bricks of the code that you produced
- Present your conclusions on data structure / trends and the final model that you used
- Q&A session with the jury

Two winners will be announced, one based on the final leaderboard, the other one based on the richness of the approach. The two winner groups will then present their work to the other students (same modalities with questions from the students)

The presentation must be as professional as possible. Here are some advices and best practices that may be useful:

- Structure your presentation: start by stating the problem that you want to solve, then present the way you tackled it, and finally describe your solution. The “story” of the presentation should be natural and easy to follow. A Jupyter notebook might be a relevant tool.
- Be concise and precise: focus on the most important messages, as you only have 5 minutes to present the work achieved.
- Be organized as a team: split up the speaking time between the team members beforehand to make it smoother
- Be honest: tell where you encountered issues or challenges
- C-suite level: you should convince both the CEO and the CTO/CDO of the company

Download instructions & submission process

How to download datasets?

You can download datasets and potential additional information at the following links:

- for_students

Where to upload you results?

You should upload and test your results at the following links:

- Hackathon Leaderboard · Streamlit (hackatonenpctheendlessline.streamlit.app)

Now is your turn!

- ✓ Find your groups of 4 people
- ✓ Put yourself in a data consultant's shoes: start structuring your approach and start exploring the data
- ✓ We will be here from 2:00 pm to 3:00 pm to validate your approach and answer your questions, as well as tomorrow from 9:00 am to 10:00 am

Nicolas

Emma

Louis