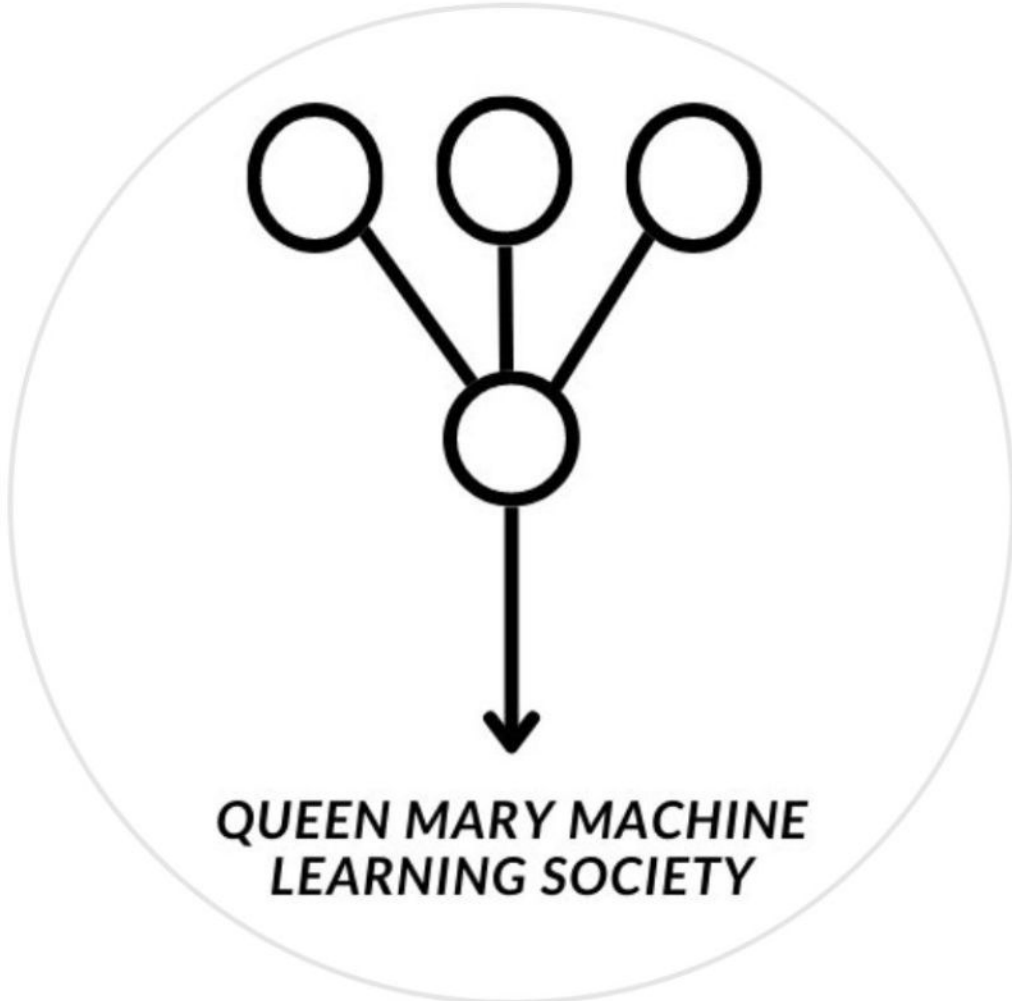
Kaggle Seasons #01



What is Kaggle

- World’s largest online platform for data science & ML competitions.
- Host to datasets, kernels (notebooks), and discussions.
- Community of 12M+ data scientists worldwide.
- A place to test your skills against others.
- Potential to earn medals, recognition, and visibility in the global data science community.

Why Kaggle

- Hands-on learning with real datasets.
- Get hands on practice on ML techniques beyond lectures.
- Build a portfolio for CV/LinkedIn.
- Work as a team → collaboration & coding skills.

What we will be doing

- Forming teams within the society
- Weekly meetups
- Monthly competitions
- Submitting results and tracking progress.
- Seasonal Team leaderboards

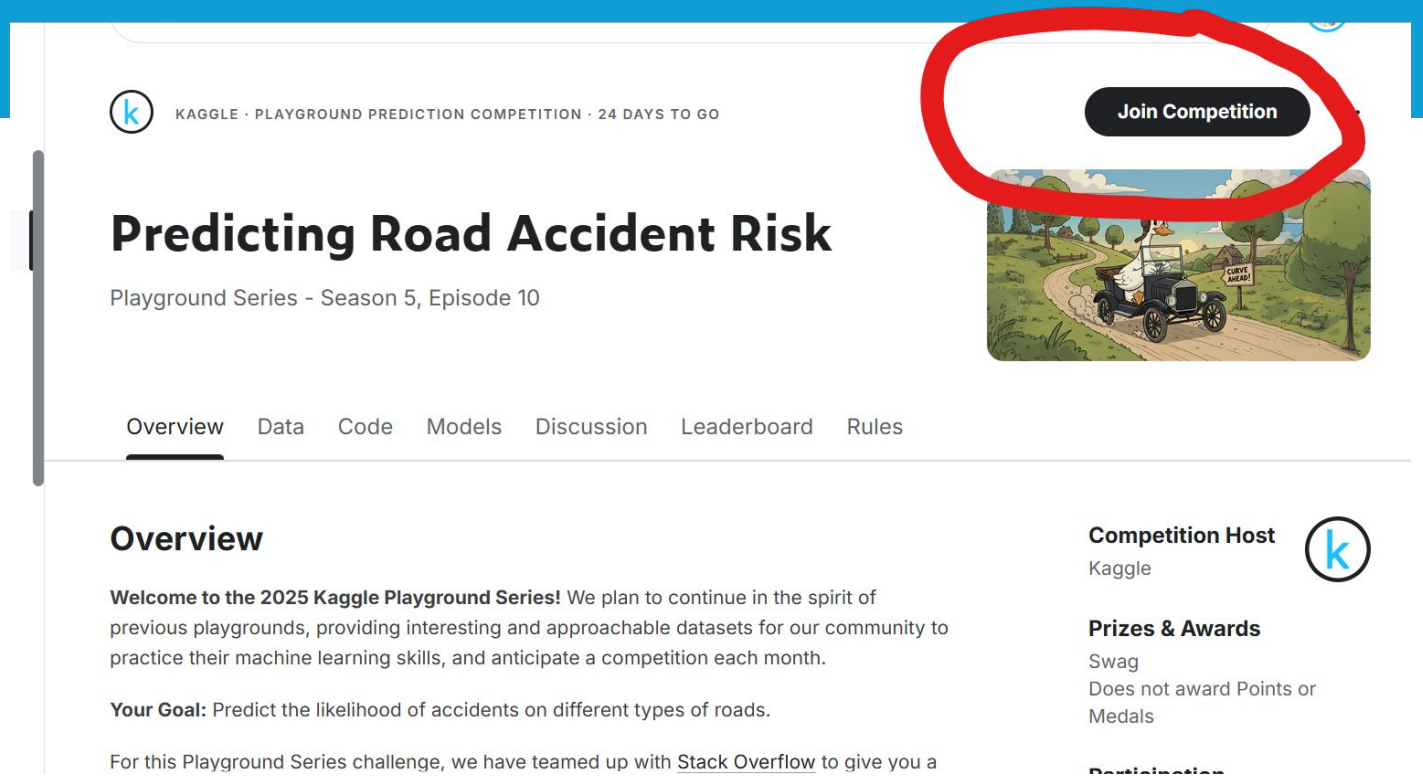
Why This Matters for You

- Boosts employability (shows applied ML experience).
- Great for CVs and interviews (“Tell me about a recent project you are proud of?).
- Exposure to real-world messy data.
- Builds confidence before coursework/research projects.

QR Code for Sign Up:

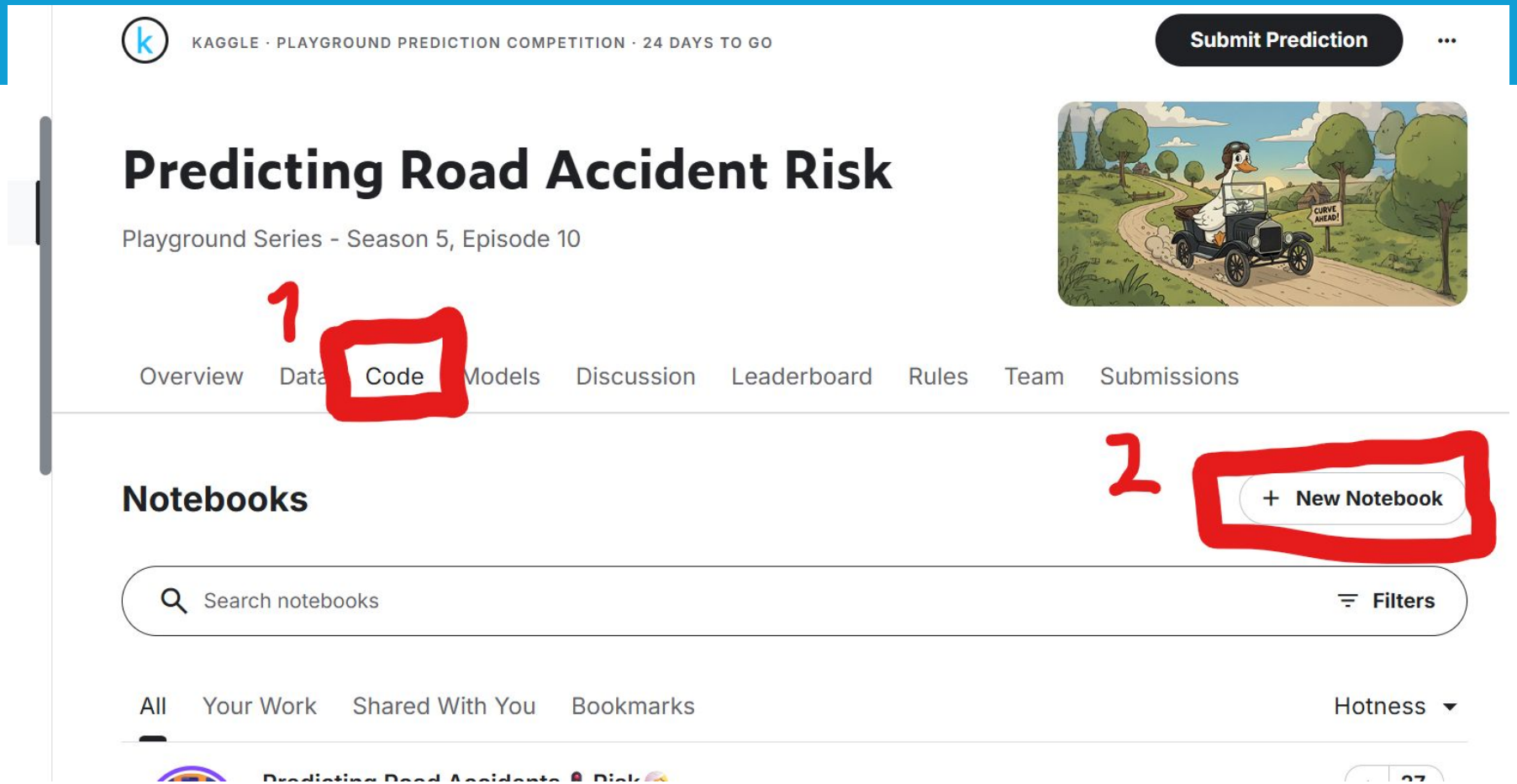
Part 2: Hands on Tutorial

Playground Series



Link to dataset: https://www.kaggle.com/competitions/playground-series-s5e10/overview
Due: October 31 2025 11:59 PM UTC

Playground Series



Run this first cell

[1]:

```
# This Python 3 environment comes with many helpful analytics libraries installe
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/doc
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets
# You can also write temporary files to /kaggle/temp/, but they won't be saved o
```

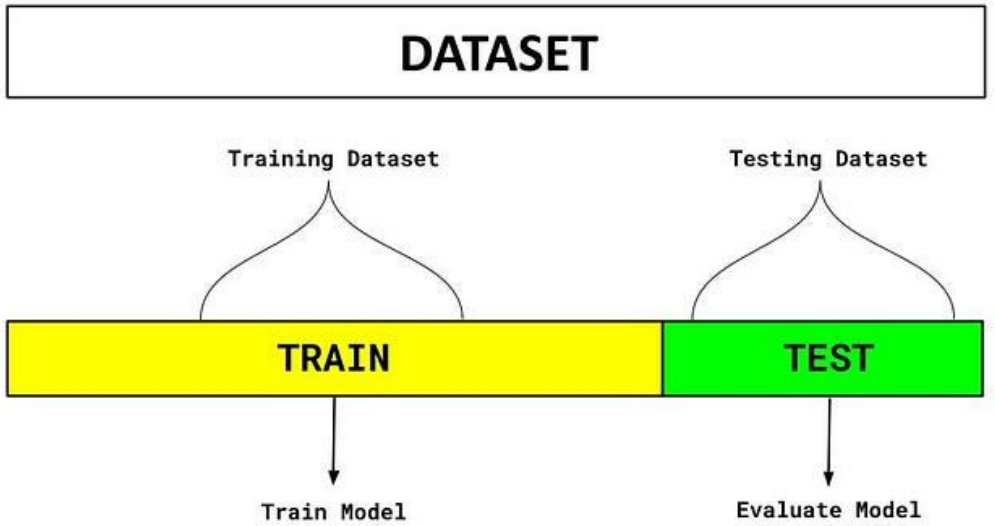
```
/kaggle/input/playground-series-s5e10/sample_submission.csv
/kaggle/input/playground-series-s5e10/train.csv
/kaggle/input/playground-series-s5e10/test.csv
```

+ Code + Markdown

Note there are 2 datasets in kaggle competitions

Training data: helps the model learn

Testing data: evaluates the model's performance on new, unseen data

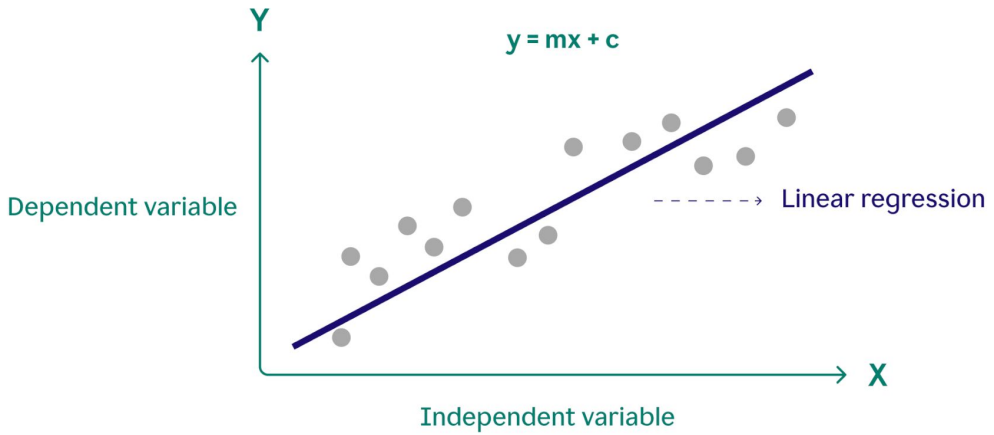


Submission.csv -> where you save your results to submit for the competition

What is X and y?

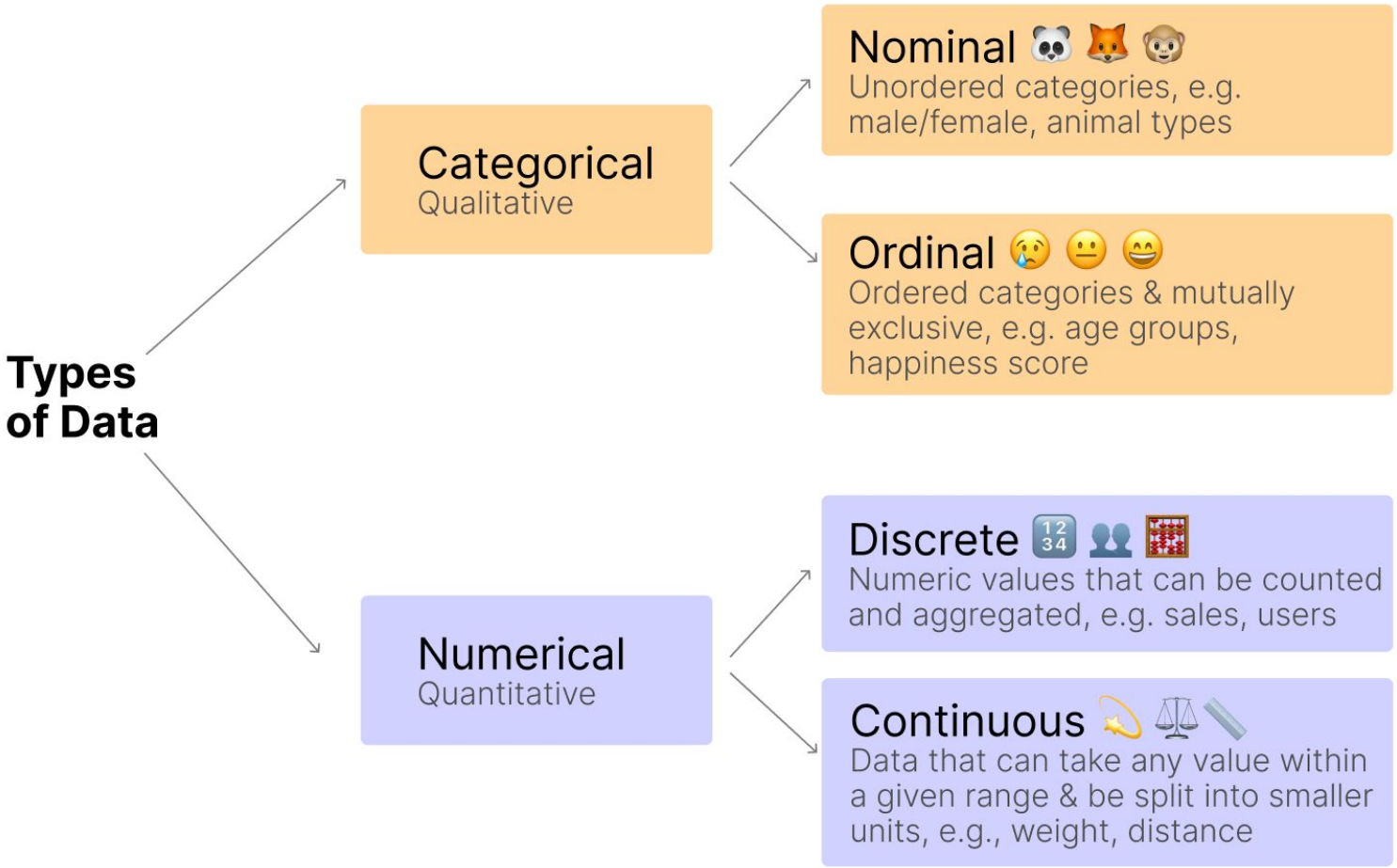
X = the input features (independent variables) used to train a model

Y = the target variable (dependent variable or output) that the model is trying to predict.



X			y
x1	x2	x3	
1	2	3	14
4	5	6	32
11	12	13	74
21	22	23	134
5	5	5	30

Categorical and Numerical data



Machine learning models are some kind of mathematical model that need numbers to work with.

Hence categorical data needs to be encoded -> turned into numbers .

(Don’t worry about this for now)