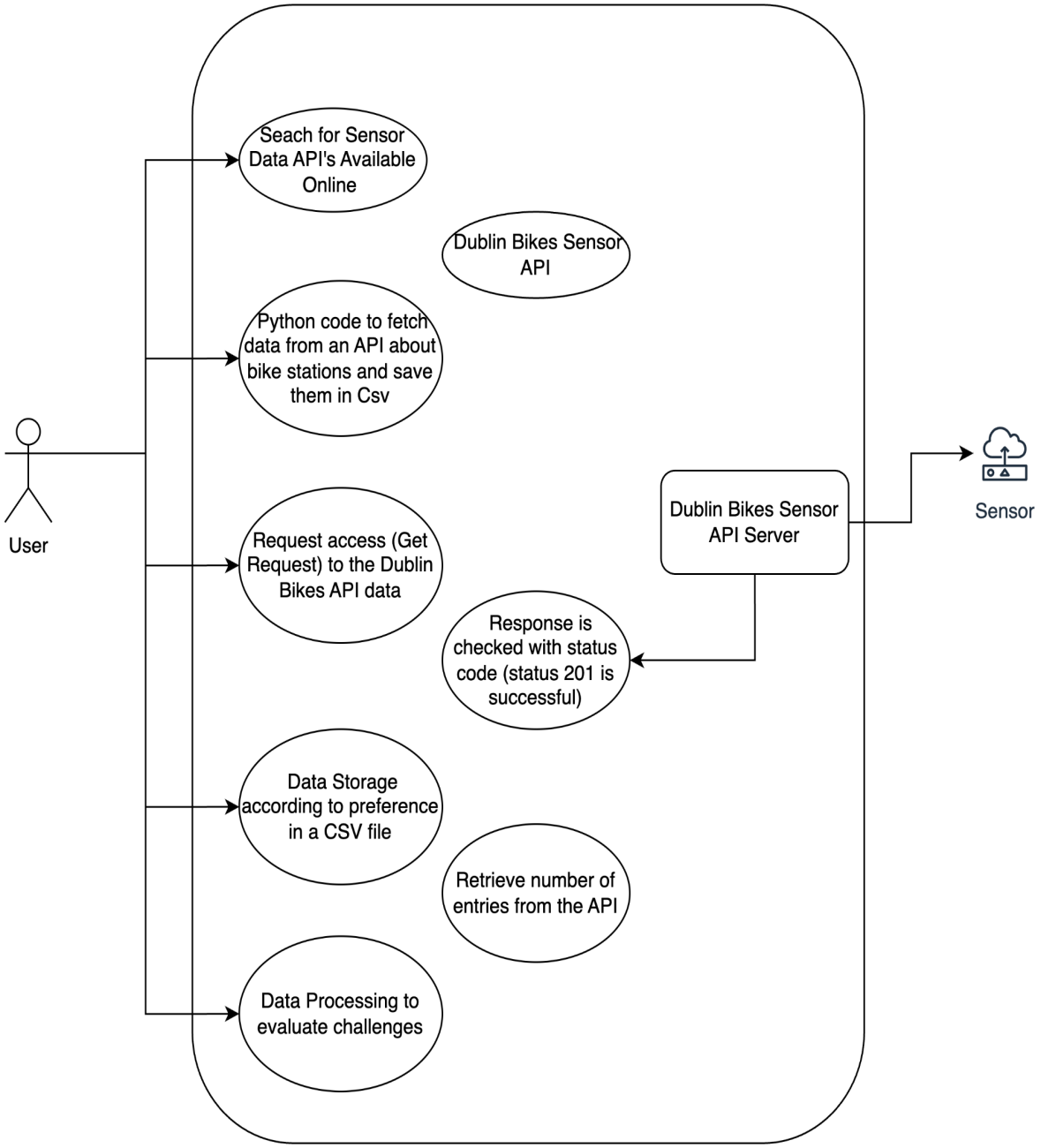
Urban Computing Week 2 Assignment
Sensor Data Collection

By: Sanat Paranjape, **Student Id**:23347997
Course Code:CS7NS4 , **Stream**: MSc CS Intelligent Systems

Task 1 - Collecting Data

Technical Diagram :



As a part of collecting data, I have used the publicly available Dublin Bikes API which uses multiple sensors on the bikes and stations associated with Dublin Bikes. The process to collect the data involves:

- **Source of Data** : As mentioned above I have used the Dublin Bikes API, which provides real-time information about the available bikes and station status. The data includes two types ‘last_snapshot’ and ‘historical’ data respectively. I have chosen last_snapshot to get the latest information and I’m retrieving 1100 entries to satisfy the number of data points.
- **Collection Method** : I have implemented python script importing requests and csv to store the API data in a csv file. HTTP GET request is sent to the API endpoint for information retrieval.
- **Data Storage** : The data retrieved from the successful API request is stored in a csv file and then further processed to assess the challenges.
- **Process of collecting data** :
    a. I made a HTTP GET request to the API endpoint at “https://data.smartdublin.ie/dublinbikes-api/last_snapshot/”.
    b. I checked the status code and it was successful (status code : 201).
    c. The JSON data was structured and was in dictionaries with one station representing a dictionary in a list.
    d. I extracted the columns and data is written to a CSV file for further discussion.

Sample of Data :

id	harvest_time	station_id	available_bike_s	bike_stands	available_bikes	banking	bonus	last_update	status	address	name	latitude	longitude
66221412	2023-10-17 20:40:02	1	27	31	4	FALSE	FALSE	2023-10-17 20:36:05	OPEN	Clarendon Row	CLARENDON ROW	53.3409	-6.2625
66221479	2023-10-17 20:40:02	2	11	20	9	FALSE	FALSE	2023-10-17 20:36:24	OPEN	Blessington Street	BLESSINGTON STREET	53.3568	-6.26814
66221500	2023-10-17 20:40:02	3	20	20	0	FALSE	FALSE	2023-10-17 20:36:47	OPEN	Bolton Street	BOLTON STREET	53.3512	-6.26986
66221440	2023-10-17 20:40:02	4	17	20	3	FALSE	FALSE	2023-10-17 20:38:30	OPEN	Greek Street	GREEK STREET	53.3469	-6.27298
66221435	2023-10-17 20:40:02	5	1	40	39	FALSE	FALSE	2023-10-17 20:38:49	OPEN	Charlemont Street	CHARLEMONT PLACE	53.3307	-6.26018
66221402	2023-10-17 20:40:02	6	17	20	3	FALSE	FALSE	2023-10-17 20:35:21	OPEN	Christchurch Place	CHRISTCHURCH PLACE	53.3434	-6.27012
66221443	2023-10-17 20:40:02	7	11	29	18	FALSE	FALSE	2023-10-17 20:38:16	OPEN	High Street	HIGH STREET	53.3436	-6.27507
66221455	2023-10-17 20:40:02	8	18	30	12	FALSE	FALSE	2023-10-17 20:39:32	OPEN	Custom House Quay	CUSTOM HOUSE QUAY	53.3479	-6.24805
66221430	2023-10-17 20:40:02	9	17	24	7	FALSE	FALSE	2023-10-17 20:31:11	OPEN	Exchequer Street	EXCHEQUER STREET	53.343	-6.26358
66221487	2023-10-17 20:40:02	10	15	16	1	TRUE	FALSE	2023-10-17 20:30:46	OPEN	Dame Street	DAME STREET	53.344	-6.2668
66221421	2023-10-17 20:40:02	11	30	30	0	FALSE	FALSE	2023-10-17 20:34:42	OPEN	Earlsfort Terrace	EARLSFORT TERRACE	53.3343	-6.2585
66221476	2023-10-17 20:40:02	12	18	20	2	FALSE	FALSE	2023-10-17 20:36:14	OPEN	Eccles Street	ECCLES STREET	53.3592	-6.26978
66221407	2023-10-17 20:40:02	13	23	30	7	FALSE	FALSE	2023-10-17 20:37:09	OPEN	Fitzwilliam Square We	FITZWILLIAM SQUARE WEST	53.3361	-6.25282
66221411	2023-10-17 20:40:02	14	18	30	12	FALSE	FALSE	2023-10-17 20:39:19	OPEN	Fownes Street Upper	FOWNES STREET UPPER	53.3446	-6.26337
66221484	2023-10-17 20:40:02	15	13	16	3	FALSE	FALSE	2023-10-17 20:39:09	OPEN	Hardwicke Street	HARDWICKE STREET	53.3555	-6.26442
66221457	2023-10-17 20:40:02	16	5	20	15	FALSE	FALSE	2023-10-17 20:34:57	OPEN	Georges Quay	GEORGES QUAY	53.3475	-6.25219
66221422	2023-10-17 20:40:02	17	18	20	2	FALSE	FALSE	2023-10-17 20:30:26	OPEN	Golden Lane	GOLDEN LANE	53.3408	-6.26773
66221403	2023-10-17 20:40:02	18	8	30	21	FALSE	FALSE	2023-10-17 20:38:33	OPEN	Grantham Street	GRANTHAM STREET	53.3341	-6.26544
66221442	2023-10-17 20:40:02	19	29	30	1	FALSE	FALSE	2023-10-17 20:39:04	OPEN	Herbert Place	HERBERT PLACE	53.3344	-6.24557
66221400	2023-10-17 20:40:02	20	29	30	1	FALSE	FALSE	2023-10-17 20:39:20	OPEN	James Street East	JAMES STREET EAST	53.3366	-6.24811
66221508	2023-10-17 20:40:02	21	11	30	19	FALSE	FALSE	2023-10-17 20:34:47	OPEN	Leinster Street South	LEINSTER STREET SOUTH	53.3422	-6.25449
66221473	2023-10-17 20:40:02	22	3	20	17	FALSE	FALSE	2023-10-17 20:37:30	OPEN	Townsend Street	TOWNSEND STREET	53.3459	-6.25461
66221413	2023-10-17 20:40:02	23	0	30	30	FALSE	FALSE	2023-10-17 20:38:07	OPEN	Custom House	CUSTOM HOUSE	53.3483	-6.25466
66221489	2023-10-17 20:40:02	24	7	20	13	FALSE	FALSE	2023-10-17 20:30:06	OPEN	Cathal Brugha Street	CATHAL BRUGHA STREET	53.3521	-6.26053
66221464	2023-10-17 20:40:02	25	30	30	0	FALSE	FALSE	2023-10-17 20:39:31	OPEN	Merrion Square East	MERRION SQUARE EAST	53.3394	-6.24655
66221482	2023-10-17 20:40:02	26	5	20	15	TRUE	FALSE	2023-10-17 20:37:13	OPEN	Merrion Square West	MERRION SQUARE WEST	53.3398	-6.25199
66221456	2023-10-17 20:40:02	27	20	20	0	FALSE	FALSE	2023-10-17 20:37:07	OPEN	Molesworth Street	MOLESWORTH STREET	53.3413	-6.25812
66221504	2023-10-17 20:40:02	28	6	30	24	FALSE	FALSE	2023-10-17 20:36:53	OPEN	Mountjoy Square Wes	MOUNTJOY SQUARE WEST	53.3563	-6.25859

I have accumulated a csv file with over 1000 points.

Discussion :

i. Identification of the relevant challenges on the dataset :

1. **Imperfection**

- **Data Quality :** The data retrieved contained some challenges with the quality. The data contained some inconsistencies because of its real time nature.

Real time data can contain some errors but processing the data efficiently will assure its quality.

- **Data Volume :** Due to real time updates to the sensor data every 5 minutes the volume of the data retrieved was more than 30000 entries which can cause problems for memory storage.
- **Imprecision :** The data may not always replicate real world situations, in this case the estimated number of available bikes may be slightly inaccurate.

2. **Correlation**

    - **Bike Availability :** The availability of bikes can have a correlation with the time of the day, for example during rush hours the bike availability will be lowest and many stations might have 0 bikes available for use.
    - **Events and Weather :** Both events and weather can affect the usage due to external factors leading to overuse or underuse of the bikes.

3. **Disparateness**

    - **Data Sources :** There are multiple data sources for Dublin Bikes sensor API but the one regularly updated is used for the assignment. The data comes from multiple sensors, the ones installed at the station, bikes , mobile applications, and the official DublinBikes Department. These multiple sensors can introduce disparateness.
    - **Historical vs last_snapshot :** Historical data may differ from the last_snapshot data as the historical data is often aggregated, structured and detailed. This will create disparateness between the data collected.

ii. Discussion as to how they could be solved:

- Imperfections can be solved using data validation and cleaning processes to handle missing values**.** There were some missing values in the data which were handled by deleting the rows having null values.
- Volume of data was reduced by incorporating about 1100 entries instead of the 30000 entries retrieved after the first request.
- Imprecision can be minimised by filtering and reducing the noise in the data available.
- Bike availability issues can be solved analysing the patterns and usage times which further can be optimised using better distribution of bikes across stations.
- Integrating weather and event data can be a starting point for resolving the impact of external factors.
- Dealing with multiple sources of data can be challenging but establishing data mapping can help in integration.
- Both historical and last_snapshot data can be analysed for trends to maintain an accurate database.

I did not make use of AI tools in the preparation of this assignment.