

Configuration Manual

MSc Research Project
Data Analytics

Vladimir Estrada Ventura
Student ID: x23427639

School of Computing
National College of Ireland

Supervisor: Abid Yaqoob

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Vladimir Estrada Ventura
Student ID:	x23427639
Programme:	Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Abid Yaqoob
Submission Due Date:	11/08/2025
Project Title:	Configuration Manual
Word Count:	508
Page Count:	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Vladimir Estrada Ventura
Date:	15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Vladimir Estrada Ventura
x23427639

1 System Overview

FáilteFinder is a modular recommender system for the Irish tourism sector, integrating NLP, KNN-based similarity search, and geospatial filtering. It ingests open datasets from the Fáilte Ireland API, processes and stores them in a PostgreSQL database, and serves interactive recommendations via a Streamlit web interface.

Main Components:

1. Data Ingestion Layer – (ingestionData.py) Downloads, cleans, and stores processed CSV files.
2. Data Store Layer – (storeProcessedData.py) Loads processed CSVs into PostgreSQL tables.
3. Keyword Extraction – (itemFrequency.py) Generates most common keywords from datasets.
4. Visualization Tool – (itemFrequencyPlot.py) Produces keyword word clouds.
5. Recommendation Engine – (recommenderEngine.py) Implements NLP vectorization and KNN retrieval.
6. Web Application – (recommender_app.py) Streamlit interface for user interaction.

2 System Requirements

2.1 Hardware:

1. CPU: Quad-core or higher
2. RAM: ≥ 12 GB
3. Storage: ≥ 10 GB free

2.2 Software:

1. OS: Windows 10+, macOS, or Linux (Ubuntu 20.04+ recommended)
2. Python 3.9+
3. PostgreSQL 14+
4. Internet connection (for API data ingestion)

2.3 Python Libraries:

- pandas
- numpy
- nltk
- gensim
- scikit-learn
- geopy
- folium
- streamlit
- streamlit-folium
- sqlalchemy
- psycopg2
- requests
- matplotlib
- wordcloud

3 Installation Steps

3.1 Python and PostgreSQL

- Download and install Python
- Install PostgreSQL and ensure psql is in PATH.

3.2 Clone or copy the project files

Place all .py files, .sql schema files, and images in the desired working directory.

3.3 Set up the database

```
psql -U postgres -f database.sql psql -U postgres -f wordsFrequencyTable.sql
```

3.4 Install dependencies

```
pip install -r requirements.txt
```

4 Configuration

4.1 Database Credentials

Update the following variables in all scripts where applicable (db_user, db_pass, db_host, db_port, db_name):

- db_user = 'postgres'
- db_pass = 1234
- db_host = 'localhost'
- db_port = '5432'
- db_name = 'recommender'

4.2 API Key

Ensure using this API key in ingestionData.py or replace it with your valid Fáilte Ireland API key:

```
headers = "Ocp-Apim-Subscription-Key": "a5f6e926adbc4b3e9f7bb40ca163e155"
```

4.3 File Structure

After running ingestion, the data/current year/ folder will store processed CSVs:

```
data/  
  2025/  
    attractions_processed.csv  
    events_processed.csv  
    accommodation_processed.csv
```

5 Execution Workflow

1. Ingest and preprocess data:

```
python ingestionData.py
```

2. Store in database

```
python storeProcessedData.py
```

3. Generate keyword frequencies

```
python itemFrequency.py
```

4. Visualize keywords (optional)

```
python itemFrequencyPlot.py
```

5. Run web application

```
streamlit run recommender_app.py
```

6 Web Application Usage

1. Open browser at the URL shown by Streamlit (e.g., <http://localhost:8501>).
2. Select accommodation status and location (map click if not booked).
3. Choose attraction and event tags (from popular list or custom).
4. Select visit month(s) or “All year”.
5. Click Get recommendations.
6. View results for attractions, events, and accommodations in tables.

7 Troubleshooting

- Database connection error
Check PostgreSQL is running and credentials are correct.
- API fetch error
Verify API key validity and internet connection.
- No recommendations returned
Ensure keywords match dataset vocabulary; run ingestion again to refresh data.
- NLTK errors
Run in Python shell:

```
import nltk
nltk.download('stopwords')
nltk.download('punkt')
```