# LGPS Data Analysis Project - Instructional File

## 1. Project Overview  
The LGPS Data Analysis Project is designed to automate the process of web scraping, data extraction, analysis, and reporting of data from Local Government Pension Scheme (LGPS) pool portfolios. The project focuses on the 'Border to Coast' and 'Brunel' pension partnership funds, generating detailed analytical insights and reports.  
  
---  
  
## 2. Project Structure  
The project is organized into the following main components:  
  
- \*\*Web Scraping Modules\*\*:   
 - `bordercoast\_data\_webscraping.py`: Scrapes documents from the Border to Coast website.  
 - `brunel\_data\_webscraping.py`: Scrapes documents from the Brunel website.  
  
- \*\*Data Cleaning and Filtering\*\*:  
 - `file\_filter.py`: Deletes unnecessary files based on predefined keywords.  
 - `complete\_file\_del.py`: Executes file deletion processes.  
  
- \*\*Data Analysis Modules\*\*:  
 - `Brunel\_LGPS\_data\_analysis.py`: Analyzes Brunel data files and generates analytical reports.  
 - `Bordertocoast\_LGPS\_data\_analysis.py`: Analyzes Border to Coast data files and generates reports.  
  
- \*\*Data Explanation Module\*\*:  
 - `data\_explaination.py`: Generates a high-level explanation of the analyzed data.  
  
- \*\*Main Execution Script\*\*:  
 - `main.py`: Orchestrates the entire workflow from web scraping to data analysis and reporting.  
  
- \*\*Setup Script\*\*:  
 - `setup.py`: Contains installation requirements and setup instructions.  
  
---  
  
## 3. Setup Instructions  
  
### Prerequisites  
- Python 3.11 or later installed  
- Install required packages using the command:  
```bash  
pip install -r requirements.txt  
```  
  
### Folder Structure  
Ensure the following folder structure is maintained:  
```  
Project\_Root/  
│  
├── Data/  
│ ├── raw\_data/  
│ │ ├── bordertocoast\_data/  
│ │ └── brunel\_data/  
│ ├── processed\_data/  
│ │ ├── bordertocoast\_data/  
│ │ └── brunel\_data/  
│ └── supporting\_files/  
│  
├── web\_scrapers/  
├── Data\_Analyzer/  
├── misc/  
└── main.py  
```  
  
---  
  
## 4. How to Run the Project  
  
1. \*\*Data Collection\*\*:  
 - Run `main.py` with `raw\_data\_generated=False` to perform web scraping and data collection.  
 ```bash  
 python main.py  
 ```  
  
2. \*\*Data Cleaning\*\*:  
 - Unnecessary files will be automatically deleted by the `complete\_file\_del.py` script.  
  
3. \*\*Data Analysis\*\*:  
 - The analysis scripts will process the cleaned data and generate Word reports with insights.  
  
4. \*\*Generating Data Explanation\*\*:  
 - The `data\_explaination.py` module will generate a summary document of the findings.  
  
---  
  
## 5. Output Files  
- Processed documents and reports will be saved in the `Data/processed\_data/` directory.  
- Supporting files (e.g., images, charts) will be saved in the `supporting\_files` subdirectories.  
  
---  
  
## 6. Troubleshooting  
  
- \*\*Common Issues\*\*:  
 - Missing dependencies: Ensure all packages in `requirements.txt` are installed.  
 - Permission errors: Run the script with appropriate permissions.  
 - Incorrect folder structure: Verify that directories are set up as outlined.  
  
---  
  
## 7. Additional Notes  
- The `setup.py` script can also be used to install the project as a package:  
```bash  
pip install .  
```  
  
- For any specific issues not covered in this document, refer to inline comments within the codebase or contact the development team.