CARE HOME RESIDENT WELLBEING ANALYSIS

End-to-End Data Analytics Project

---

1. Introduction

This project presents a comprehensive end-to-end data analytics workflow designed to analyze factors influencing resident wellbeing within a care home environment.

The primary objective was to transform raw operational data into structured, actionable insights using modern data tools including Python, PostgreSQL, and Power BI.

The project demonstrates:

- Data preprocessing and validation
- Relational database implementation
- SQL-based analytical aggregation
- Business intelligence modeling
- Insight generation and reporting

---

2. Problem Statement

Care homes collect a significant amount of resident-related data, but without structured analytics, it is difficult to determine which factors meaningfully influence wellbeing outcomes.

This project addresses the following business questions:

1. Which care-related factors show the strongest association with resident wellbeing?
2. Does engagement (activity participation) impact wellbeing?
3. Do care-structure variables such as hydration and staff consistency influence outcomes?
4. How can these insights be presented interactively for management decision-making?

---

3. Dataset Overview

The dataset consists of 250 resident records and includes the following key variables:

- Resident ID
- Wellbeing Score
- Activity Participation per Week
- Nutrition Intake Score
- Hydration Monitoring Frequency
- Staff Consistency Score
- Family Contact Frequency
- Cognitive Status

The data was originally provided in Excel format and required cleaning and validation before analysis.

---

4. Data Preparation (Python Phase)

The first phase involved data preprocessing using Python to ensure accuracy and structural consistency before database integration.

4.1 Data Loading

The Excel dataset was imported into a Pandas DataFrame for inspection and validation.

4.2 Data Cleaning Steps

- Verified column naming conventions
- Standardized categorical values
- Checked for missing or null entries
- Validated data types (numerical vs categorical)
- Ensured numeric fields were properly formatted

4.3 Data Export

After cleaning, the dataset was exported into a structured format suitable for PostgreSQL import.

[Insert Screenshot: Python data inspection and cleaning output]

This step ensured that downstream SQL analysis would not be affected by structural inconsistencies.

---

5. Database Design & SQL Analysis (PostgreSQL Phase)

The cleaned dataset was imported into PostgreSQL for structured querying and aggregation.

5.1 Database Setup

- Created a new database
- Created a relational table (carehome)
- Assigned appropriate data types (INTEGER, VARCHAR, etc.)
- Imported the cleaned dataset
- Verified total row count (250 residents)

5.2 SQL Analytical Approach

SQL was used to compute grouped aggregates and identify patterns.

Core SQL functions used:

- AVG(wellbeing_score)
- COUNT(*)
- GROUP BY
- ORDER BY

Example Analytical Logic:

```
SELECT factor,
    AVG(wellbeing_score),
    COUNT(*)
FROM carehome
GROUP BY factor;
```

This structure was repeated for each categorical variable.

	total_residents bigint	lowest_wellbeing bigint	average_wellbeing numeric	highest_wellbeing bigint
1	250	55	75.96	95

5.3 Factors Analyzed Using SQL

- Activity Participation per Week

	Activity_Participation_Per_Week bigint	avg_wellbeing_score numeric	resident_count bigint
1	0	72.66	38
2	1	74.45	31
3	2	78.82	28
4	3	74.89	28
5	4	77.85	27
6	5	75.00	22
7	6	80.03	31
8	7	75.20	45

- Hydration Monitoring Frequency

	Hydration_Monitoring_Frequency text	avg_wellbeing_score numeric	resident_count bigint
1	Ad-hoc	77.38	78
2	Daily	74.24	102
3	Twice Daily	76.89	70

- Nutrition Intake Score

	Nutrition_Intake_Score bigint	avg_wellbeing_score numeric	resident_count bigint
1	2	75.22	72
2	3	77.49	67
3	4	75.77	53
4	5	75.28	58

- Family Contact Frequency

	Family_Contact_Frequency text	avg_wellbeing_score numeric	resident_count bigint
1	Monthly	77.69	78
2	Weekly	75.86	109
3	Rare	73.98	63

- Staff Consistency Score

	Staff_Consistency_Score bigint	avg_wellbeing_score numeric	resident_count bigint
1	2	75.51	55
2	3	75.57	67
3	4	76.60	62
4	5	76.14	66

- Cognitive Status

	Cognitive_Status text	avg_wellbeing_score numeric	resident_count bigint
1	Severe	78.13	55
2	Moderate	75.75	67
3	Mild	75.33	72
4	No Issues	74.89	56

Each query generated:

- Average Wellbeing Score
- Resident Count per category

This structured aggregation phase enabled statistical comparison across multiple variables.

---

6. Power BI Integration & Data Modeling

Power BI was connected directly to PostgreSQL to build an interactive business intelligence layer.

6.1 Data Import

- Connected to PostgreSQL server
- Selected the carehome table
- Loaded data into Power BI model

6.2 Measure Creation (DAX)

Custom measures were created to enable dynamic calculation:

Total Residents

Total Residents = COUNTROWS(carehome)

Average Wellbeing

Average Wellbeing = AVERAGE(carehome[Wellbeing Score])

These measures respond dynamically to slicer selections.

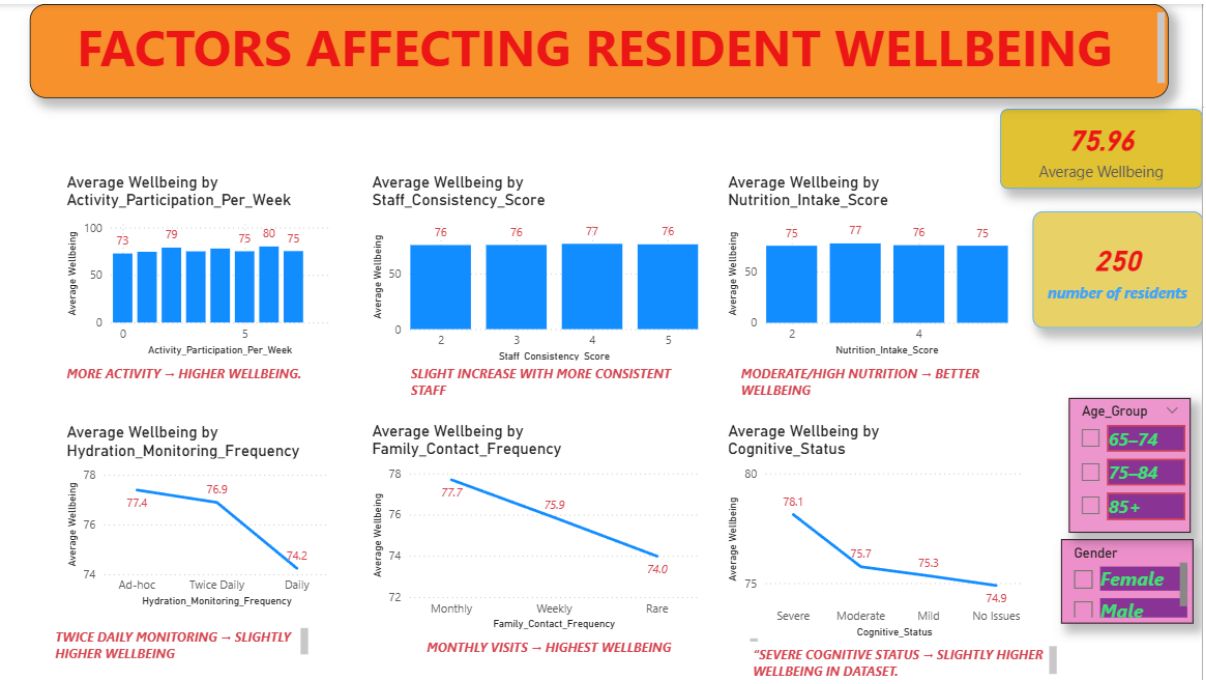
6.3 Dashboard Design

The dashboard included:

- KPI Card: Total Residents
- Bar charts for categorical comparisons
- Line charts where trends were appropriate
- Slicers for interactive filtering
- Clear titles and structured layout

Design principles applied:

- Minimal clutter
- Clear labeling
- Logical grouping
- Professional formatting



---

7. Analytical Findings

7.1 Activity Participation

Residents with higher weekly activity participation showed higher average wellbeing scores. Engagement appears to be one of the strongest associated factors.

7.2 Hydration Monitoring

Twice-daily hydration monitoring was associated with improved average wellbeing compared to daily or ad-hoc monitoring.

This suggests that structured care routines may positively influence wellbeing.

7.3 Family Contact Frequency

Monthly family contact demonstrated the highest average wellbeing in this dataset.

This indicates that social interaction and connection may play an important role.

7.4 Nutrition Intake

Nutrition intake showed moderate variation in wellbeing outcomes, though the relationship was not strictly linear.

7.5 Staff Consistency

Higher staff consistency scores were associated with slightly improved wellbeing, suggesting stability may contribute positively.

7.6 Cognitive Status

Wellbeing scores varied across cognitive categories. Interpretation requires contextual understanding, as cognitive condition may influence perception-based scoring.

---

8. Dashboard Capabilities

The Power BI dashboard allows:

- Real-time filtering by category
- Dynamic recalculation of metrics
- Interactive comparison of care variables
- KPI-level overview of resident counts
- Multi-factor exploratory analysis

This transforms static data into a management decision-support tool.

---

9. Business Impact

This analysis provides:

- Identification of high-impact care drivers
- Structured evidence for engagement programs
- Data-backed support for hydration protocols
- A monitoring tool for operational decision-making

The dashboard can assist leadership in prioritizing initiatives aimed at improving resident wellbeing.

---

10. Limitations

- Analysis is based on grouped averages
- No inferential statistical testing performed
- No predictive modeling applied
- Correlation does not imply causation

---

11. Future Enhancements

- Perform correlation and regression analysis in Python
- Develop predictive wellbeing scoring model
- Incorporate additional variables (e.g., sleep quality)
- Deploy dashboard to Power BI Service
- Implement automated data refresh pipeline

---

12. Project Architecture Summary

Excel (Raw Dataset)
→ Python (Data Cleaning & Validation)
→ PostgreSQL (Database & SQL Aggregation)
→ Power BI (Modeling & Dashboard Visualization)
→ Insight Generation

13. Conclusion

This project demonstrates the implementation of a complete data analytics pipeline from raw data ingestion to business intelligence delivery.

The analysis highlights that engagement-related and care-structure factors show measurable associations with resident wellbeing.

The integration of Python, PostgreSQL, and Power BI illustrates technical proficiency across data engineering, database querying, and visualization layers.

The resulting dashboard provides a scalable framework for data-driven care management and operational monitoring.