Power BI project – Medokanja Healthcare Data

Table of Contents

1. Project Name : Medokanja

This project focuses on improving healthcare systems by analyzing wait lists, which record individuals awaiting medical treatment, consultations, procedures, or diagnostic tests due to resource limitations, scheduling constraints, or high demand.

The name Medokanja combines "Medo" (medical) and "Kanja" (patient in Japanese).

2. Overall Objectives

This project aims to analyze patient waiting lists to improve healthcare systems.

The dashboard provides:

- Historical insights (from 2018 to 2021) to identify patterns, enabling better planning and reduced waiting times.
- Detailed analysis of specialties with longer wait times, identifying the medical areas where patients face the most delays or where the waiting list is the longest, allowing targeted improvements.
- Insights into age groups with extended waits, showing which patient demographics experience delays, enabling strategies to address their specific needs.

3. Data Overview

The project utilizes publicly available healthcare data focused on patient waiting lists. This dataset is divided into two categories:

- Inpatients are individuals who require hospital admission due to the complexity of their treatment or the need for extended monitoring and care. Examples include surgeries requiring post-operative observation or treatment for serious conditions like pneumonia.
- Outpatients receive care that does not necessitate overnight hospitalization. Common examples include routine check-ups, diagnostic tests, or minor procedures.

4. High-Level Data Study

a. Data Sources

The project leverages three main data sources:

1. **Inpatient Data**: Stored in the Inpatient folder, with four CSV files representing data from 2018 to 2021. File names follow the format IN_WL_year (e.g., IN_WL_2018.csv). The four CSV files have the same number of columns and identical column headers.

2. **Outpatient Data**: Stored in the Outpatient folder, with four CSV files representing data from 2018 to 2021. File names follow the format Op_WL_year (e.g., Op_WL_2018.csv). The four CSV files have the same number of columns and identical column headers.

Inpatient Data

Archive_Date	Specialty_HIPE	Specialty_Name	Case_Type	Adult_Child	Age_Profile	Time_Bands	Total
31-01-2018	0	Small Volume Specialities	Inpatient	Child	0-15	6-9 Months	1
31-01-2018	0	Small Volume Specialities	Inpatient	Child	16-64	9-12 Months	1
31-01-2018	400	Endocrinology	Day Case	Child	0-15	3-6 Months	1
31-01-2018	400	Endocrinology	Day Case	Child	0-15	12-15 Months	1
31-01-2018	600	Otolaryngology (ENT)	Day Case	Child	0-15	0-3 Months	14

Outpatient Data

Archive_Date	Specialty_HIPE	Specialty	Adult_Child	Age_Profile	Time_Bands	Total
31-01-2018	100	Cardiology	Child	0-15	0-3 Months	167
31-01-2018	100	Cardiology	Child	0-15	3-6 Months	157
31-01-2018	100	Cardiology	Child	0-15	6-9 Months	106
31-01-2018	100	Cardiology	Child	0-15	9-12 Months	85
31-01-2018	100	Cardiology	Child	0-15	12-15 Months	83

3. **Specialty Mapping Data**: A single CSV file named Mapping_Specialty.csv that maps specialties to broader specialty groups, enabling the creation of categories (buckets) for improved visualization and analysis.

Specialty	Specialty Group
Haematology	Blood
Metabolic Medicine	Blood
Paed Haematology	Blood
Paed Metabolic Medicine	Blood
Orthopaedics	Bones
Paed Orthopaedic	Bones
Rheumatology	Bones
Clinical Neurophysiology	Brain
Developmental Paediatrics	Brain
Neurology	Brain
Neurosurgery	Brain
Paediatric Neurology	Brain
Paediatric Neurosurgery	Brain
Paed Oncology	Cancer
Paediatrics	Child Care
Plastic Surgery	Cosmetic
Maxillo-Facial	Dental

b. Data Dictionary

Inpatient and Outpatient Data:

Column name	Description	Data type	Example
Archive_Date	The date when the record was archived.	DATE	2021-12-31
Specialty_HIPE	The HIPE (Hospital In-Patient Enquiry) code representing the specialty.	INTEGER	101
Specialty_Name / Specialty	The name of the medical specialty	TEXT	Cardiology
Case_Type	The type of case (e.g., "Inpatient" or "Day Case").	TEXT	Inpatient

Adult_Child	Indicates whether the patient is an adult or a child.	TEXT	Adult
Age_Profile	Categorizes patients by age groups.	TEXT	16-45
Time_Bands	Categorizes patient waiting times.	TEXT	3-6 months
Total	The total number of patients for the specified category.	TEXT	TEXT

We noticed the following differences between the inpatient and outpatient datasets:

- Specialty_Name (in inpatient) and Specialty (in outpatient) are two fields serving the same purpose. It would be better to standardize this column across both files and name it Specialty for consistency.
- Case_Type is only present in the inpatient data. To maintain consistency, we can add this column to the outpatient data and populate it with the value "Outpatient".

The Archive_Date column in both files is stored as text, not as a date. To convert it correctly, right-click the Archive_Date column, select Change Type > Using Locale, choose Date, and set the locale to French (France) for accurate conversion.

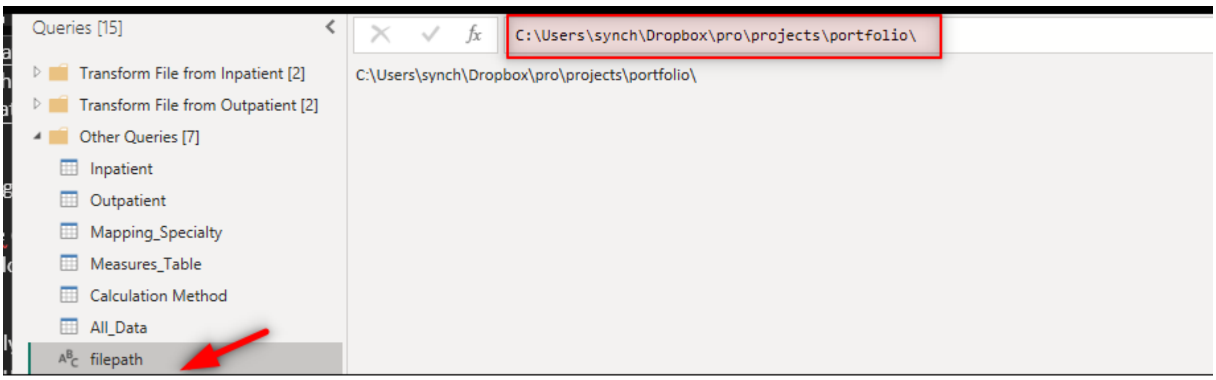
Specialty Mapping Data

Column name	Description	Data type	Example
Specialty	The specific medical specialty.	TEXT	Cardiology
Specialty_Group	The broader category to which the specialty belongs.	TEXT	Heart

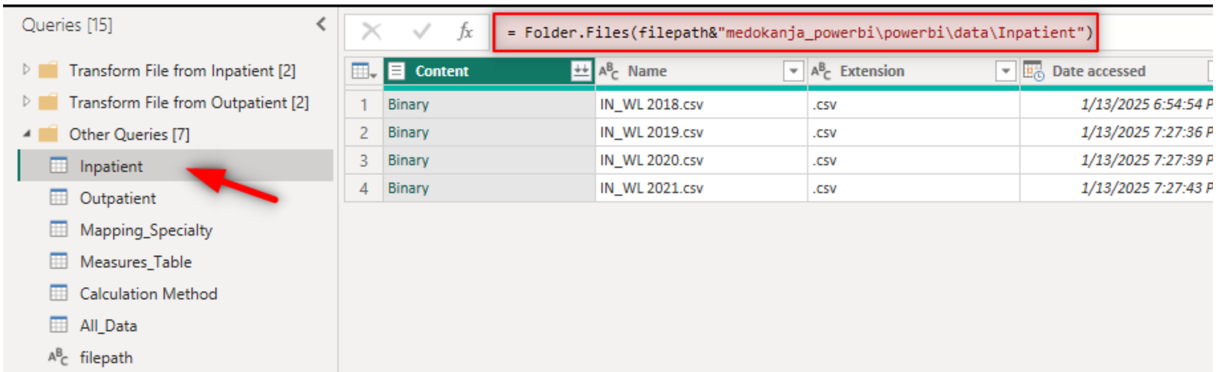
5. Load Data

The project dynamically uses the filepath parameter to locate data files across three main sources. This approach ensures flexibility and portability, allowing users to configure the file path based on their local system.

The filepath parameter is configured in Power Query to dynamically adjust the folder path. Users only need to update the filepath value to point to their local directory (e.g., C:\Users\YourName\Documents\), and all queries will automatically adapt to locate the data files.



Here is an example of the Inpatient Source:



6. Data Consolidation

1. **Combine** the four CSV files in each folder (Inpatient and Outpatient) into two tables: Inpatients and Outpatients.
2. **Standardize Columns**:
    - Rename the Specialty_Name and Specialty columns in both files to Specialty.
    - Add a new column, Case_Type, to the Outpatients table and populate it with the value "Outpatient".
3. **Append** the Inpatients and Outpatients tables into a new table named All_Data, while retaining the raw Inpatients and Outpatients tables for reference.

7. Relationship

Establish a relationship between the All_Data table and the Mapping_Specialty table using the common Specialty column, with a cardinality of One-to-Many (1:N) from the perspective of Mapping_Specialty to All_Data.

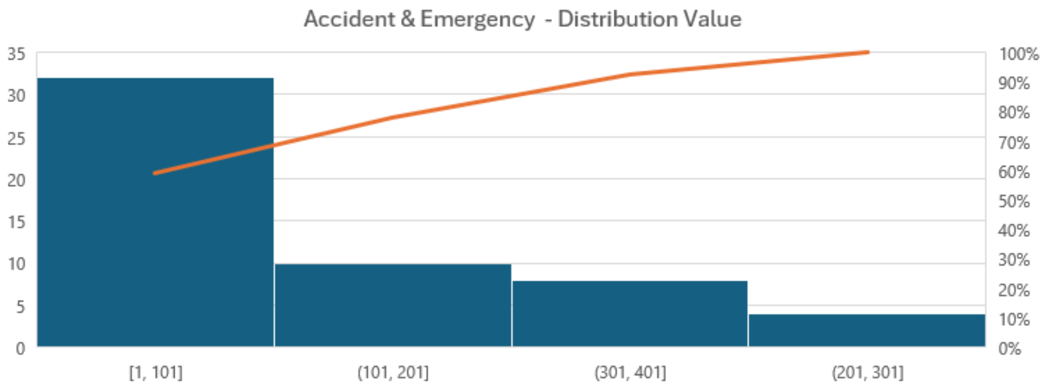
8. Data Analysis

a. Data Distribution, Average vs. Median

The dataset analysis indicates that the data is not normally distributed, as the mean, median are not close to each other. For example:

- Dermatology: Average = 99, Median = 35
- Accident & Emergency: Average = 111, Median = 69.5

Specialty	Average	Median
Accident & Emergency	111.19	69.50
Anaesthetics	27.33	10.00
Breast Surgery	68.68	5.00
Cardiology	56.66	28.00
Cardio-Thoracic Surgery	8.55	5.00
Chemical Pathology	5.83	4.00
Child/Adolescent Psychiatry	7.38	3.00
Clinical (Medical) Genetics	68.78	29.00
Clinical Immunology	44.14	11.00
Clinical Neurophysiology	47.02	21.00
Clinical Pharmacology	5.11	3.00
Dental Surgery	11.81	3.00
Dermatology	98.62	35.00
Developmental Paediatrics	46.52	11.00
Diabetes Mellitus	16.16	10.00
Endocrinology	43.38	19.00
Gastro-Enterology	46.37	11.00
Gastro-Intestinal Surgery	11.60	7.00
General Medicine	53.56	15.00
General Surgery	41.22	10.00
Geriatric Medicine	20.82	7.00
Gynaecology	39.76	8.00
Haematology	27.73	15.00
Hepato-Biliary Surgery	5.99	3.00



This distribution is not normal; it's right skewed with most values concentrated in the lower range.

This highlights the presence of **outliers** and skewed distribution.

- Average: Sensitive to outliers, which can inflate or distort the central tendency.
- Median: Provides a more reliable central value in the presence of skewed data or extreme values.

Using median allows for better decision-making by focusing on the "middle ground" rather than being influenced by outliers.

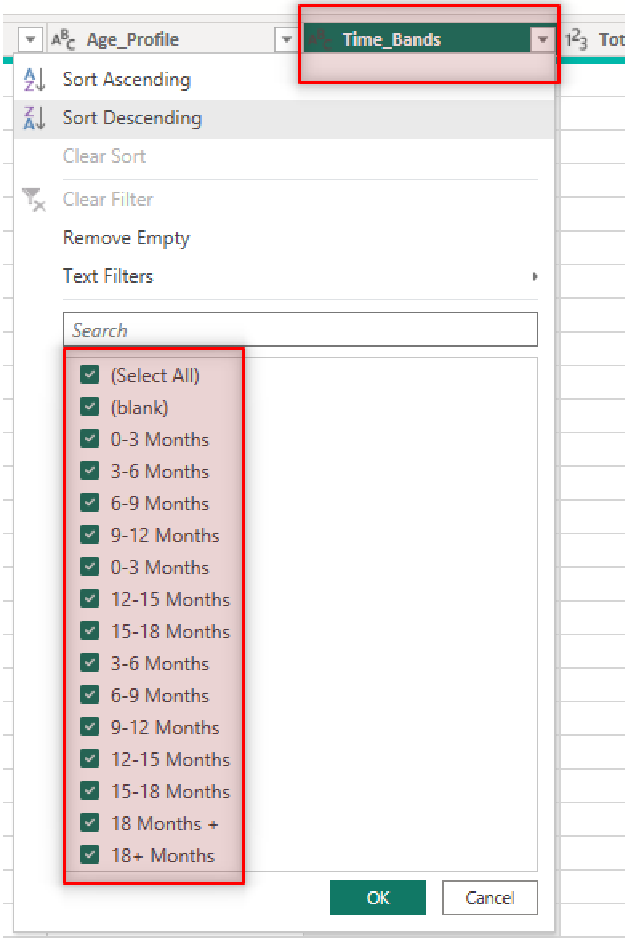
For example, Dermatology's average (99) is inflated due to a few extreme values, whereas the median (35) offers a more accurate representation of typical waitlist times.

Including both average and median in the dashboard provides stakeholders with a dynamic view:

- Stakeholders can toggle between the metrics to understand how outliers impact their analysis.
- This flexibility helps in addressing different analytical needs and scenarios.

b. Duplicate Data

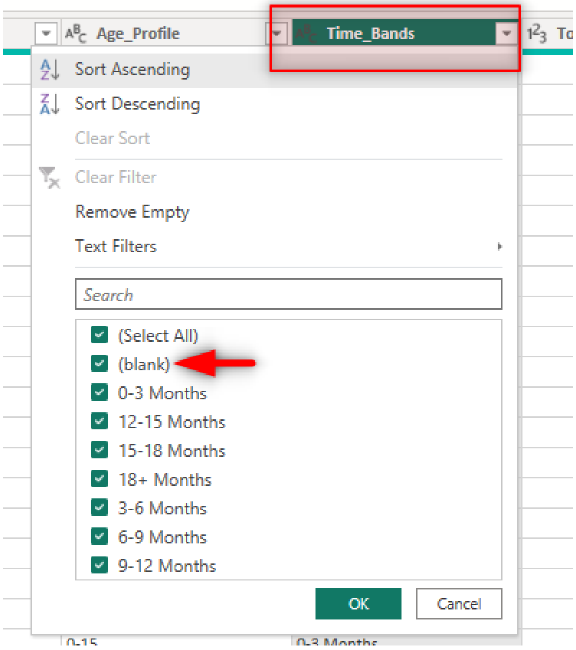
We notice duplicated data in Time_Bands column (All_Data table) and this can cause issues in analysis and visualizations.

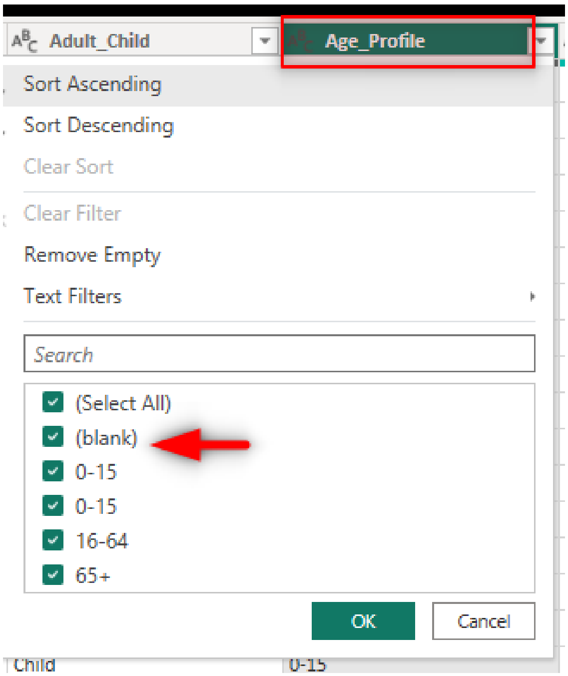


We notice the same in Age_Profile column.

The solution is to standardize these values using transformations like **Replace Values** or **Trim** to ensure consistency.

c. Missing Values





Missing data in Time_Bands and Age_Profile appears as blank values, which can disrupt analysis. The solution is to replace blank values with a placeholder like "No Input" to maintain consistency and ensure the data is usable in visualizations.

9. Design Blueprint

- Summary Page for an overall view of the data.
- A Detailed Page for granular analysis and deeper exploration.

Hex code to use in this project

a. Summary Page

CY vs PY

Latest month's wait list for the current year (CY) vs. same month's wait list for the previous year (PY)

Measure	Formula
LatestMonthWL	CALCULATE(SUM(All_Data[Total]), All_Data[Archive_Date] = MAX(All_Data[Archive_Date])) + 0

	MAX(All_Data[Archive_Date]): Gets the latest date in the Archive_Date column.
PYSameMonthWL	CALCULATE(SUM(All_Data[Total]), All_Data[Archive_Date] = EDATE(MAX(All_Data[Archive_Date]), -12)) + 0 EDATE(..., -12): Shifts the date back by 12 months to the same month in the previous year.

Dynamic Measure: Average vs. Median

Create a dynamic measure to toggle between average and median. This was done using a slicer with options ("Average" and "Median") and a DAX measure using the SWITCH function. The dynamic measure changes based on the slicer selection, allowing users to switch between displaying the average and median values in the same chart. This approach provides flexibility for analyzing data with outliers.

Case Type Split

A donut chart is created to display the split by case type (e.g., inpatient, outpatient, day case). The chart uses a dynamic measure for average/median waitlist, allowing toggling between the two metrics. The case type field is added to the legend, and the chart dynamically updates based on user selections, providing an intuitive view of the waitlist distribution across case types.

Waitlist by Time Bands & Age

A stacked column chart is created to display the average/median waitlist by time bands and age profile.

A dynamic measure is used to toggle between average and median values, allowing the chart to update based on user selection.

The time bands are added to the x-axis, age profile to the legend, and the dynamic measure to the y-axis.

Issues such as duplicate or inconsistent values (e.g., spaces or typos) in time bands and age profiles are cleaned during the transformation step to ensure accurate visualization.

Top 5 Specialties

A multi-row card is created to display the top 5 specialties based on the waitlist metric (average/median). A filter is applied to show only the top 5 specialties using the DAX measure for average/median waitlist. This allows stakeholders to focus on the key specialties with the highest wait times.

Monthly trend – Inpatients vs. outpatients

A line chart is created to display the monthly trend of inpatients vs. outpatients.

- The archive date was added to the x-axis, and the total waitlist to the y-axis.
- The case type field was used in the legend to differentiate between inpatients, outpatients, and day cases.
- Separate line charts were created for inpatients/day cases and outpatients to provide a clearer view, and filters were applied to isolate the relevant case types for each chart.

Dynamic title

A dynamic title is created to indicate whether average or median was selected.

- This was achieved using a DAX measure with the SWITCH function.
- The measure checks the user's selection from the toggle (slicer) and updates the title accordingly, displaying "Average Waitlist" or "Median Waitlist."
- The dynamic title ensures clarity for users by explicitly showing which metric is currently displayed in the charts.

b. Detailed Page

A detail page to display granular-level data is created. Key points include:

- A matrix visualization was used to show detailed data with rows for attributes like specialties and columns for case types (e.g., inpatient, outpatient).
- Filters for time bands, age profiles, and other dimensions were added on the left for better interactivity.

10. Testing

If the specialty selected in the filter has no outpatient waitlist data (i.e., the sum of waitlist for outpatient = 0), the corresponding chart becomes empty. Similarly, if the specialty has no day case or inpatient waitlist data (i.e., the sum of waitlist for day case and inpatient = 0), the related chart also becomes empty. To address this, two measures were created to dynamically check for blank results based on the filter context and display a message in place of the chart. This ensures users are informed about the absence of data instead of encountering blank or empty visuals.

Measure	**Formula**
NoDataLeft	IF(ISBLANK(CALCULATE(sum(All_Data[Total]),All_Data[Case_Type]<>"Outpatient")), "No data for selected specialty", "")
NoDataRight	IF(ISBLANK(CALCULATE(sum(All_Data[Total]),All_Data[Case_Type]="Outpatient")), "No data for selected specialty", "")

Note that the same field ("Outpatient") has been used for simplicity.

11. Interactive Views

- Navigation buttons were added to switch between the summary page and the detail page for seamless user experience.
- Advanced formatting and interactivity, like synced filters, were implemented to ensure consistency across pages.

12. Global Analysis

a. Key Insights vs Action Plan

Key Insights	Action Plan
The total waitlist is extremely high at **709K**, with outpatients contributing **74%** of this number.	Expand outpatient capacity by extending service hours, and redistributing resources from lower-pressure case types.
Pediatric specialties (e.g., Orthopedics, Urology) show severe backlogs, especially in the 0–3 month band.	Increase pediatric specialist availability, create dedicated pediatric service units, and prioritize pediatric cases in the 0–3 month band.
General specialties like Accident & Emergency, ENT, and Dermatology exhibit high median waitlists.	Strengthen resource allocation for high-pressure specialties, especially in Day Case and Outpatient categories.
Most patients are concentrated in the 0–3 month band, especially in the 16–64 age group, across all case types.	Allocate resources to quickly address short-term delays, focusing on the 16–64 age group, and use fast-track processes for critical cases to prevent long-term delays.
The 18+ month wait list is still high because the shorter wait periods (like 0-3 months, 3-6 months, etc.) are not being handled properly or in a timely manner.	Tackle shorter wait periods by reallocating resources and speeding up processing to prevent cases from reaching the 18+ month category.

b. Critical Metrics to Monitor

- Total waitlist numbers by case type (e.g., Outpatient, Inpatient, Day Case).
- Median waitlist by time band, with a focus on the 0–3 month band.
- Specialty-specific waitlists, particularly for Accident & Emergency, ENT, Dermatology, Orthopedics, and Pediatric Urology.
- Age group distribution trends, analyzing how different age groups are represented in the waitlist data over time.