

MSc in Computing (Enterprise Software Systems)

**Business Intelligence**

**Assignment, December 2021**

Vibin Varghese

**(Student Id: 20096057)**

*Submitted Date*

8rd December, 2021

**Table of Contents**

Contents

[1 Introduction 4](#_Toc89876973)

[1.1 Objective 4](#_Toc89876974)

[1.2 Target Organisation/User 4](#_Toc89876975)

[1.3 Covid-19 Vaccination Dataset 4](#_Toc89876976)

[1.4 User KPI’s defined and explanation 5](#_Toc89876977)

[2 Data Preparation 6](#_Toc89876978)

[2.1 Exploration 6](#_Toc89876979)

[2.2 Cleaning 7](#_Toc89876980)

[2.3 Combining data 8](#_Toc89876981)

[2.4 Data flow 9](#_Toc89876982)

[3 Visualisation 11](#_Toc89876983)

[3.1 Approach – Exploratory 11](#_Toc89876984)

[3.2 Alignment 11](#_Toc89876985)

[3.3 Data representation choices and Presentation 11](#_Toc89876986)

[3.4 Dashboard 13](#_Toc89876987)

[4 Limitations And Conclusion 15](#_Toc89876988)

[4.1 Limitation 15](#_Toc89876989)

[4.2 Conclusion 15](#_Toc89876990)

[5 References 16](#_Toc89876991)

**List of Figures**

[**Fig. 1.** Tableau prep builder – Flow 6](#_Toc89877975)

[**Fig. 2.** Filter and Change Role 7](#_Toc89877976)

[**Fig. 3.** Column Split 8](#_Toc89877977)

[**Fig. 4.** Pivot Data 8](#_Toc89877978)

[**Fig. 5.** Applied Join Data 8](#_Toc89877979)

[**Fig. 6.** Union Data 9](#_Toc89877980)

[**Fig. 7.** Age Group wise Data Flow 9](#_Toc89877981)

[**Fig. 8.** Manufacture Type – Data Flow 9](#_Toc89877982)

[**Fig. 9.** Vaccination Data – Data Flow 10](#_Toc89877983)

[**Fig. 10.** Color Palette 11](#_Toc89877984)

[**Fig. 11.** Horizontal Bar Chart – Vaccination per hundred people – Age Group 12](#_Toc89877985)

[**Fig. 12.** Stacked Area chart – Vaccination per hundred – Monthly/Quarterly 12](#_Toc89877986)

[**Fig. 13.** Map Chart – Ireland vs Other European Union 12](#_Toc89877987)

[**Fig. 14.** Donut Chart – Vaccination Type Administered 13](#_Toc89877988)

[**Fig. 15.** Ireland Vaccination Dashboard – Year 2021 14](#_Toc89877989)

Introduction

Objective

The main objective is to prepare a Vaccination dashboard based on the data available in the source GitHub (covid-19-data) using the Tableau prep builder and Tableau desktop.

Target Organisation/User

Main goal of the Dashboard is to find the number of peoples vaccinated in Ireland and What type of vaccines are given to them over the year 2021. Govt. officials and the public who regularly check the total vaccination administrated in Ireland are the target organisation and the users of this dashboard. By this dash board these group can understand the types of vaccine used in how much percentage and vaccinated people per hundred. They can able to compare the total percentage of people vaccinated in Ireland and other remaining European Union Countries.

Covid-19 Vaccination Dataset

In the middle of Covid-19 pandemic, by the beginning of the year 2021 all the countries started giving vaccination to the people from the age above 80. And this process continues in the current date, since vaccination is the only source to reduce the death rate.

Many originations started collecting the raw vaccination data and make it available for the public to use. For this vaccination dash board, I have used the data set from “Our World in Data” GitHub Account [[1]](#footnote-1), specifically the vaccination data. World population data also used in the dashboard; the original source data used by ‘Our World in Data’ team [[2]](#footnote-2). Three data set are mainly used in the dashboard-Location, Manufacture Wise Vaccination and Vaccination also the World Population.

Few key points of the data set

* **Total Vaccinations**: Total number of doses administered. For vaccines that require multiple doses, each individual dose is counted, i.e. For first dose, metric added 1 and again adds 1 for second dose; Total will be 2.
* **Total Vaccinations per hundred**:  Total vaccinations per 100 people in the total population of the country.
* **People Vaccinated**: Total number of people who received at least one vaccine dose. Even the person receives two doses, metric will be set as one for that person.
* **People vaccinated per hundred**: People vaccinated per 100 people in the total population of the country.
* **People Fully Vaccinated**: Total number of people who received all doses prescribed by the vaccination protocol. If a person receives the first dose of a 2-dose vaccine, this metric stays the same. If they receive the second dose, the metric goes up by 1.
* **People Fully Vaccinated Per Hundred**: People fully Vaccinated per 100 people in the total population of the country.

*Note that, data set day wise data is a running total values, which means it’s an incremental value.*

User KPI’s defined and explanation

A Key Performance Indicator (KPI) is a measurable value and is an integral part of analytics today. The main audience of this dashboard will be the Govt. Officials and general public. On a current trend on Covid-19, officials and public have two questions when come to vaccination

1. How quickly the people are getting vaccinated in the country?
2. What type of vaccine is given to the people?

Every public is looking to get vaccinated as soon as possible and prefer to know vaccine type in use. For the govt. officials, the vaccination rate and the availability of the vaccine in the country is more important. The answers to the above question can give them a better insight of Covid Vaccination.

KPI which I map to this dashboard is to make sure the questions are answered and the users can have the insight of the vaccination data. Dashboard is focusing on the Country **Ireland** for the **year 2021** when the vaccination is started.

Two KPI for the dashboard is

1. Vaccinated People per hundred
2. Type of Vaccines

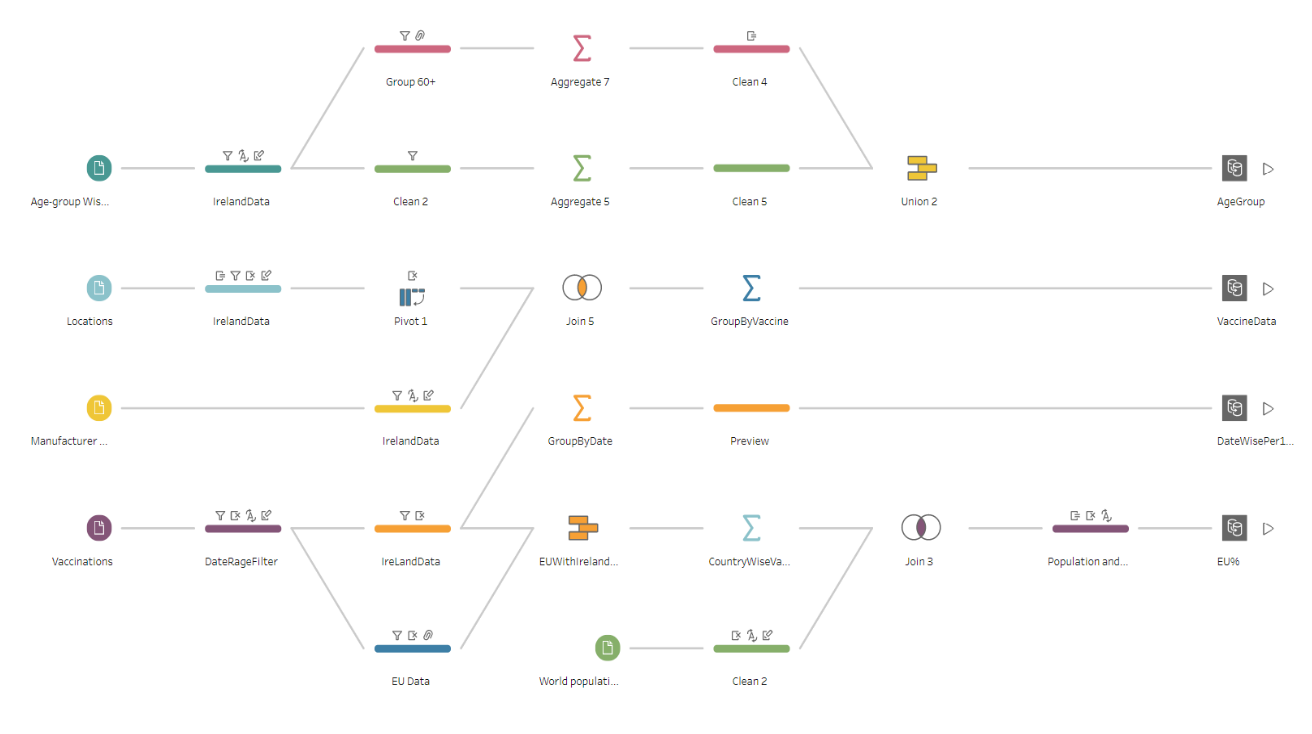
**Vaccinated People:** Dashboard will show the details of peoples vaccinated per hundred both in Quarterly wise, Month wise and Age Group wise. Total Number of Vaccination, People fully vaccinated and Total people vaccinated are shown here. The details of these are mentioned in the dataset, shows how it is calculated. Dashboard also shows the percentage of Total people vaccinated in the other European Union to compare this with country Ireland.

**Type of Vaccines:** In a country like Ireland, Govt. is providing different types of vaccines manufactured by different labs. The dashboard will give a better understanding the type of vaccines used in the country, with total number of vaccines administrated and the percentage of vaccine type used.

Data Preparation

As mentioned in the dataset details, data is been captured from the ‘Our World in Data’ GitHub account, where they regularly update the data. The data set used is till mid of November 2021. For preparing the dashboard, I have used

1. Vaccination Data
2. Location Data
3. Age-group Wise Vaccination
4. Manufacturer Wise Vaccination
5. World population

Most of the data is correct for the country Ireland. And in tableau prep builder, I have modified few of the fields, remove not utilized field and done necessary grouping and joins based on the date.

**Fig. 1.** Tableau prep builder – Flow

Exploration

Five data set is used in the flow as mentioned before. Detail exploration of each data set in given below.

1. **Vaccination Data** – This data set have country wise, date wise data of vaccination; Which include the total vaccination, people vaccinated, fully vaccinated, booster dose data, per hundred data, per million and so on. Most of the data in this dataset is incremental value, i.e., the running total day wise. There are few null values in few countries but the for the KPI of this dashboard used only the Ireland and other European Union countries, which is pretty good.
2. **Location Data** – Source of data for each country and the last date of data fetch from the source is mention in this dataset. This includes a column which shows the vaccine type used in the country.
3. **Age-group Wise Vaccination** – The vaccination dataset does not mention the age group of the people, which will be use full for the dashboard. This data set show age group data in day wise for people vaccinated and fully vaccinated. But the data set have data issue for the age group above 60. This is mentioned in the limitation of the dash board.
4. **Manufacturer Wise Vaccination** – Other than location data set, all other dataset is not having the vaccine type details. This dataset contains County wise, day wise total vaccination with type of vaccine. This is also an incremental dataset.
5. **World population**- GitHub account provides the source of world population they have used in preparing the dataset. Same data source is used in the dashboard. This is having the data of population country wise for the year 2021, with source of data mentioned.

Cleaning

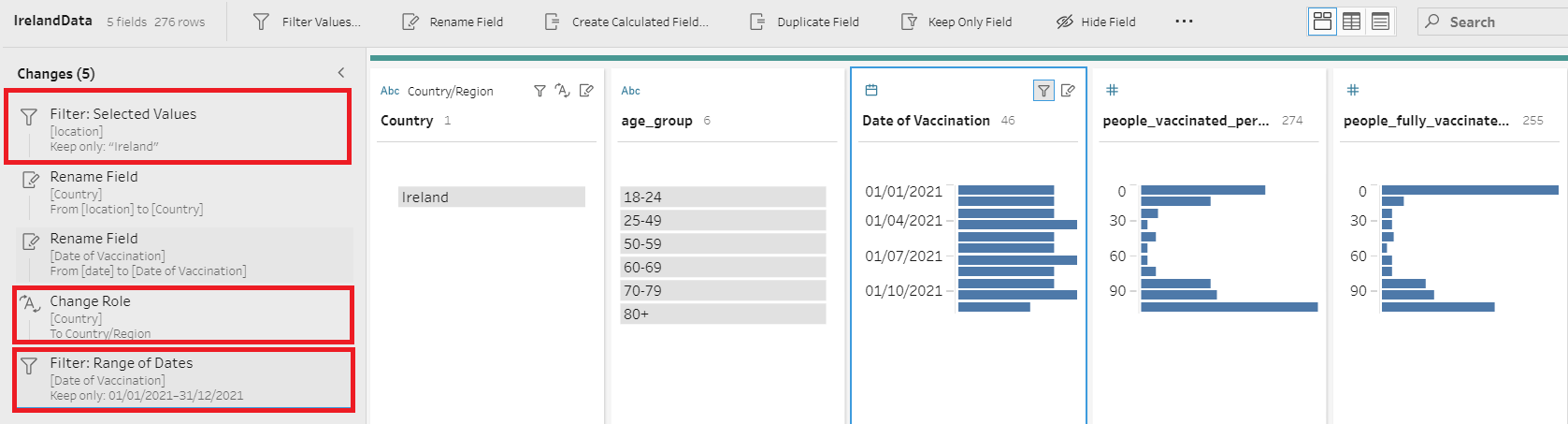
The data set is been cleaned in different stages of the flow based on the output required. All the data set with a filter of Country and the date range from January 1st to November 30th 2021. There are some entries before this date in Vaccination data, which is not clear and have null data in most of the places.

**Filter:** Country = Ireland

Range of date = 01/01/2021 – 30-11-2021

For the dataset to get data of European Union countries, Country is filtered in later stage in the above (Fig. 1) flow of Vaccinations.

**Columns Rename/Removed:** Few fields in the data set have changed the column name for better understanding and unnecessary data has been removed from the dataset to reduce the content of output. I have kept few fields extra in the data set for the future use, when new set of data representation is required in the future.

**Change Role:** In all the data set location column is changed to country and Type of the Column is changed to region, this is done majorly for the Map, where countries are used.

**Fig. 2.** Filter and Change Role

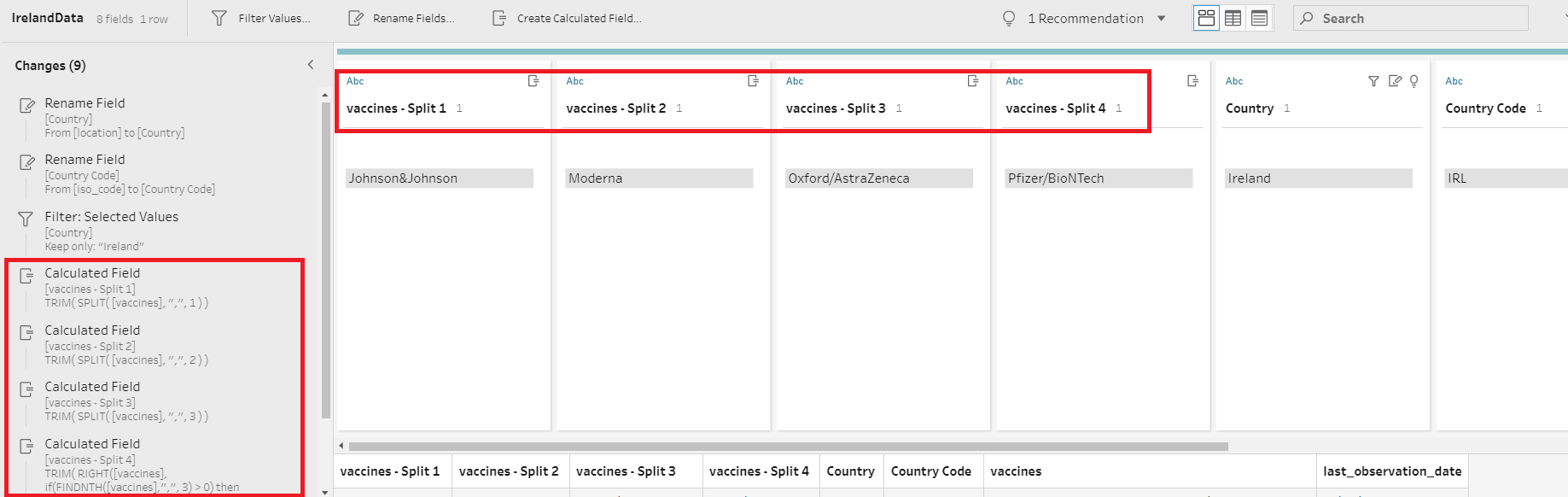
**Group Values:** In the age group wise data set have issue (explained in the limitation) in the data. The data above the age of 60-69, 70-79 and 80+ is group to 60+. This value is used later for calculation and in the dashboard.

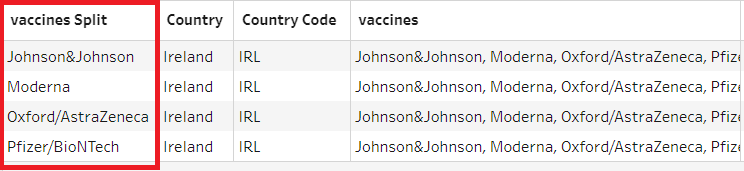
**Calculated Filed:** Few calculations are required in the data set, to get the sum of values, max of values and Average of data. Other than that, values in the data set and calculation are Round to two decimal points for better understanding in the dashboard.

**Calculation Field – Split Data** – Have used the column split for one of the data sets. Split of data is done using comma. This is done for the location field where vaccine type is mentioned in one single column.

**Pivot** - Location dataset have vaccine details in one column, which is split to multiple columns is pivoted to rows for joining to the manufacture dataset.

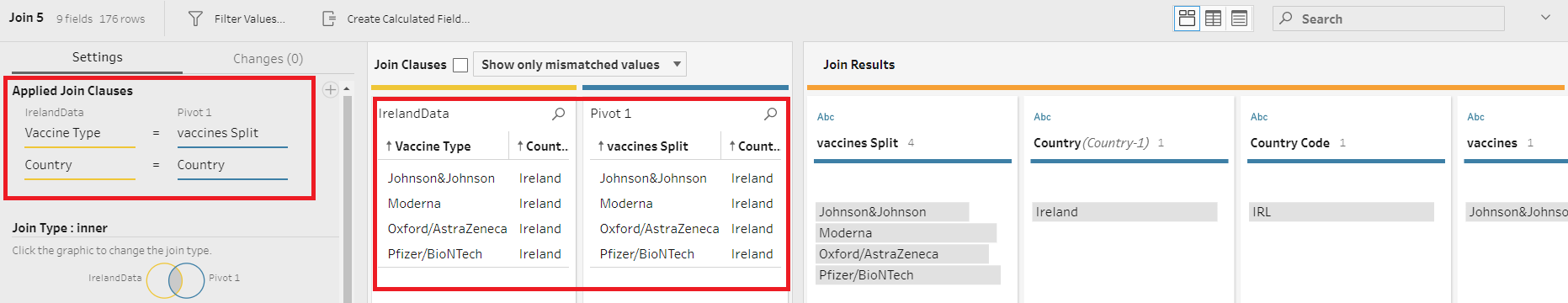
**Grouping** – Age Group data set, grouping is done to get the Max and Average values Age Group wise. For manufacture dataset, data is combined based on the type of vaccines used, and for vaccination data, is grouped by the date for one output and by country wise for the other data set.

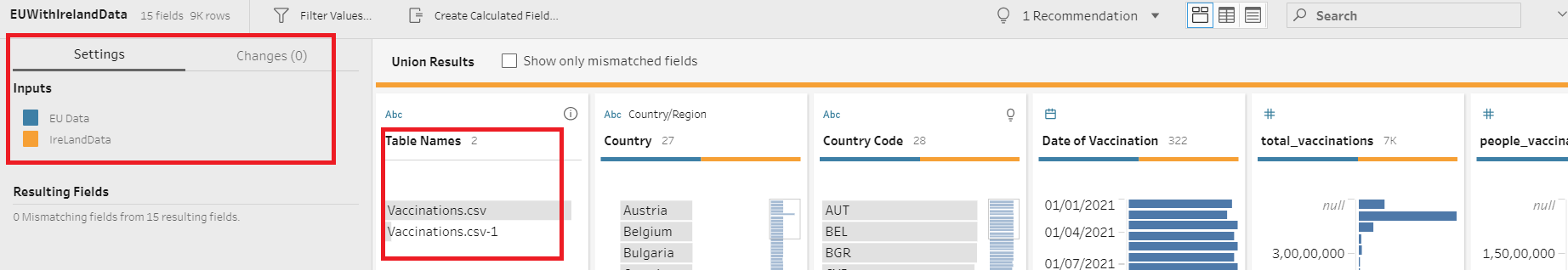
**Fig. 3.** Column Split

**Fig. 4.** Pivot Data

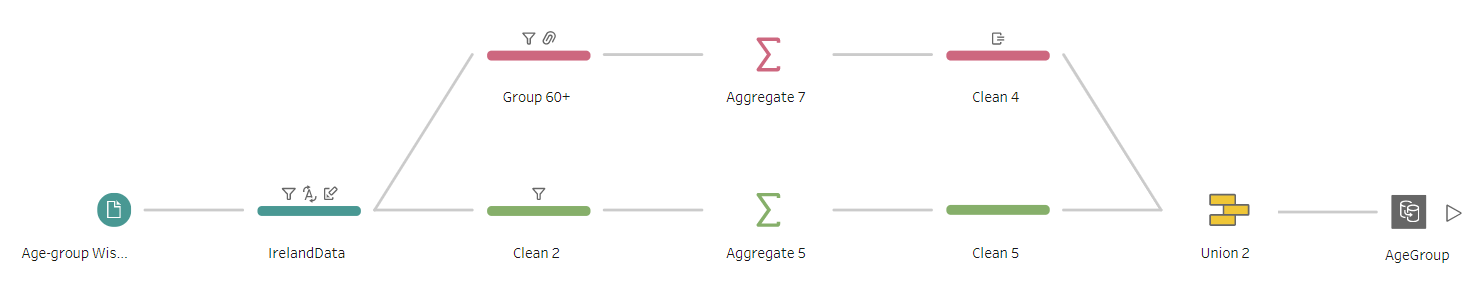
Combining data

**Joins** – Two joins are used in the prep builder, first one for joining the Ireland data with other European union data. And next one is to get the vaccine in location match with the manufacture dataset, so that no extra data will come from the manufacture data set.

**Fig. 5.** Applied Join Data

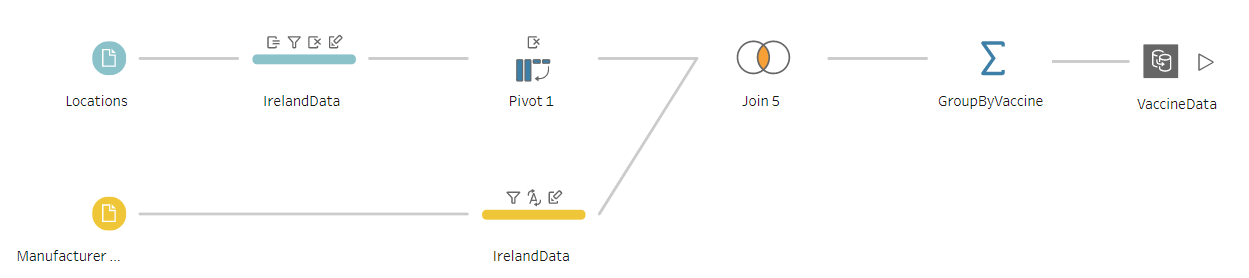
**Union** – Data from the vaccination have two branch which will filter Ireland and other EU countries separately for two outputs. This is combined together for getting the data for map representation.

**Fig. 6.** Union Data

Data flow

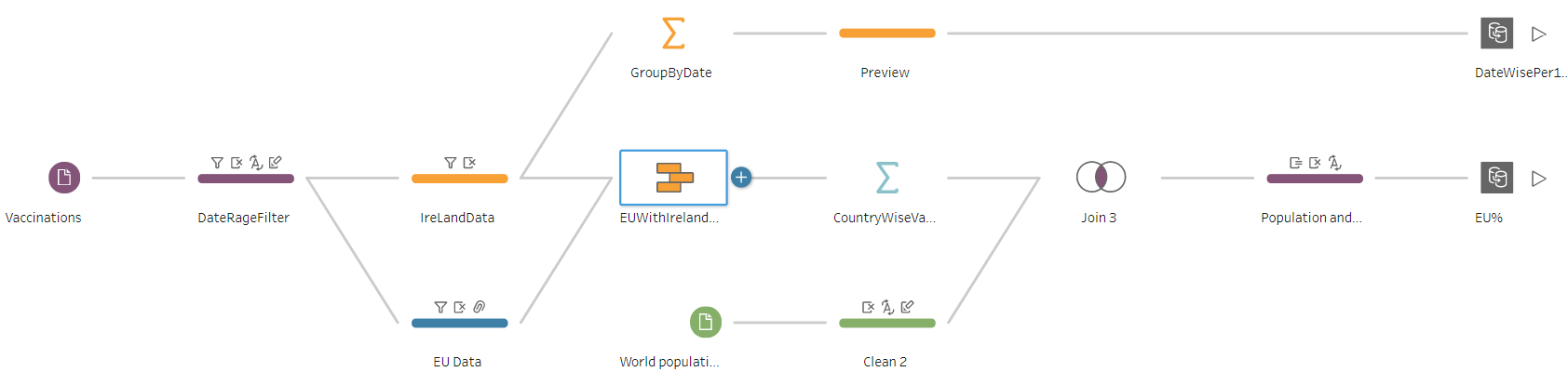
Three set of data flow is there in the prep builder for this dashboard.

**Fig. 7.** Age Group wise Data Flow

Age GroupWise (**Fig. 7**), the dataset of this is cleaned first with Country filter and renamed two fields and change the type of Location to Region. Clean 2 process will filter only three set of age group and on next step, this is grouped and find the max value from it. Group60+ clean process will filter only one age group 60+, and change the remaining age group above to 60+ group and on the next step average of the data is take. Both aggregate 7,5 will be union to get the output of this data set.

**Fig. 8.** Manufacture Type – Data Flow

Next data flow is to get the Manufacture wise vaccine type data (**Fig. 8**), The location data set is used here to cross check the type of vaccine listed in manufacture data set is same as the location. Location data is filter for country and split with comma for vaccine column, which will give four columns of vaccine. And using the pivot data is changed from column to row. This will join with the manufacture data set by country and vaccine type and then group to get the max vaccination for each type.



**Fig. 9.** Vaccination Data – Data Flow

For vaccination date wise data and European union data (**Fig. 9**), Vaccination and World population data set is used. Vaccination data is filtered with date and on the next clean steps, one creates only Ireland data and other for European union data. Ireland data is later grouped by date and output is generated. EU data and Ireland data is union and group based on country. The output of this is join with world population using country. From the joined data, Percentage value is calculated and get the output for the Map chart.

Visualisation

This section will explain the visualisation process of the data we have cleaned using prep builder. To visualise the data, we have used the Tableau Desktop.

Approach – Exploratory

For the dashboard, we have used five data set and giving the users an option to explore the data using this dash board. The dashboard will help to explore the following

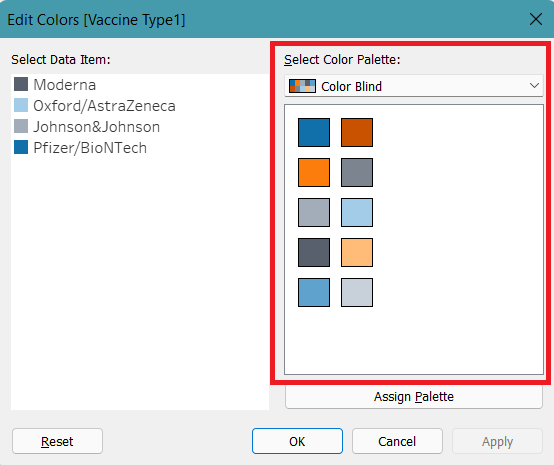
* Vaccination administrated per hundred (Age Group and Month wise/quarterly).
* Comparison of Ireland vaccination and other European Union Countries.
* Percentage of different type of vaccines given to people in Ireland.

Alignment

* Dashboard Size: Generic Desktop (1366 X 768)
* Layout: Floating – All elements are floating, including the legends and filters.
* Grid: 20 px – Used to move and resize floating elements pixel by pixel

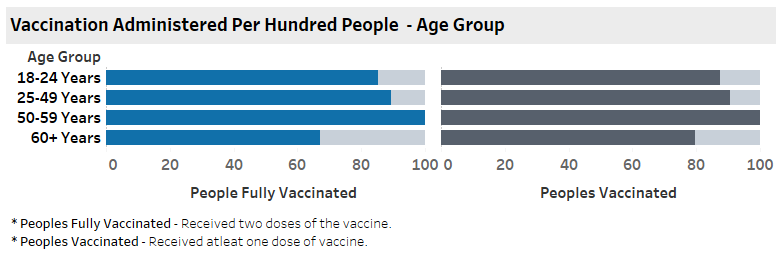
Data representation choices and Presentation

**Color choice:** I have used ‘Color Blind’ palette in the entire dashboard. In the Tableau color blind-friendly palette under both deuteranope and protanope simulation blue and black shades will be almost same, since the dashboard is also be used by the general public as mention in the targeted organization/users this palette is used.

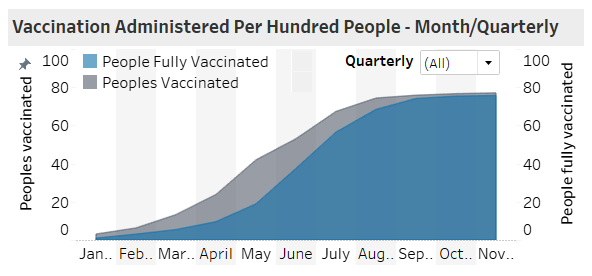


**Fig. 10.** Color Palette

**Bar Chart Horizontal:** Age group wise vaccination administrated by hundred people is represented in a horizontal bar chart. Reason to use this chart, Its Comparing among age group over one value per item and have only four categories to list out. Also because of long label name.

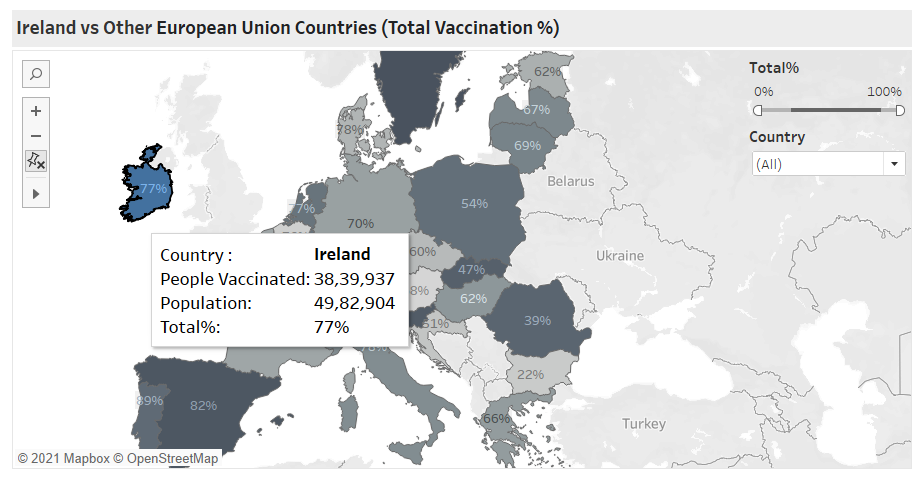


**Fig. 11.** Horizontal Bar Chart – Vaccination per hundred people – Age Group

**Stacked Area chart:** Vaccination administrated by hundred people months wise/Quarterly, I have used this Stacked area chart. It’s a Composition of values, which is changing over the different date range and can show the relative and absolute differences of vaccination done for fully vaccinated and peoples totally vaccinated.

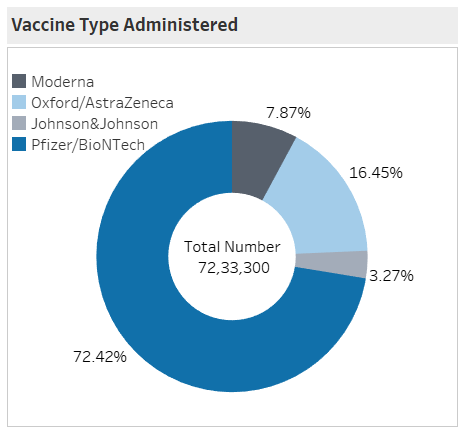
**Fig. 12.** Stacked Area chart – Vaccination per hundred – Monthly/Quarterly

*Note: The color legends for people fully vaccinated and people vaccinated for the above two charts are identical; this is for the users' better understanding.*

**Map chart:** To show the comparison of Ireland and other European Union countries, Map chart will show categories across geographical regions with percentage of Vaccination done in the country.

**Fig. 13.** Map Chart – Ireland vs Other European Union

**Donut chart:** I have converted the pie chart to donut chart for this dashboard. Set of vaccine type can be divided into multiple segments in proportion with the related values, by Donut chart I can able to show the total number of vaccines so that user can instantly compare it with the segment values.

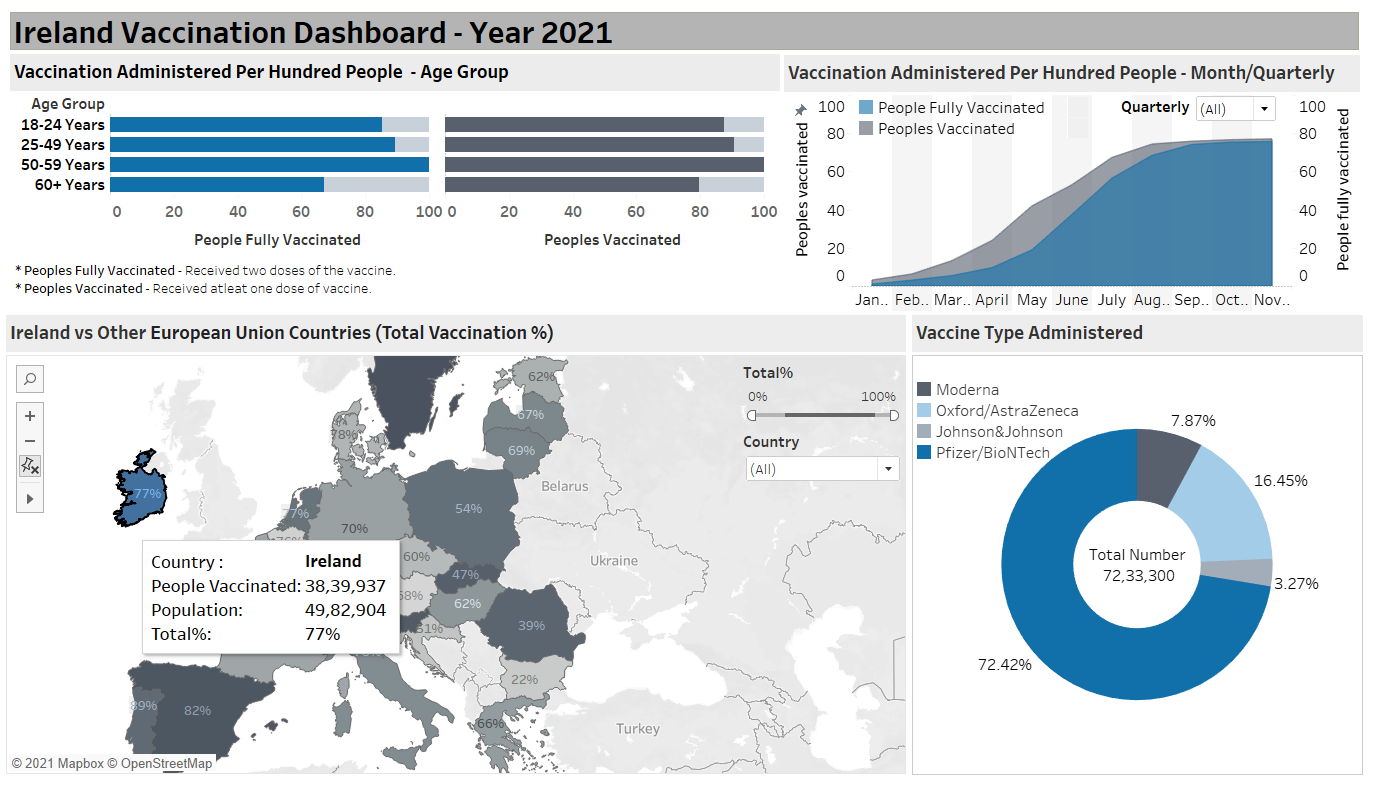


**Fig. 14.** Donut Chart – Vaccination Type Administered

Dashboard

Final dashboard is shown in the Fig. 15. The dash board have four charts

1. **Vaccination Administered per hundred people** - Age group: This chart will show the fully vaccinated people and the total people vaccinated on different four age group. The value will be less than or equal to 100, since it’s a per hundred rates.
2. **Vaccination Administered per hundred people** – Month/Quarterly: This chart will show the fully vaccinated people and the total people vaccinated by month wise, filter option is given to get data quarterly also. User can select multiple quarter while filtering the data.
3. **Ireland vs Other European Countries (Total Vaccination %)** – Chart is to show the percentage pf people vaccinated in these countries. Ireland is highlighted, since it’s an Ireland dashboard. Two set of filtration given in this chart, Range filter 0% to 100% and Country filter for checking and comparing multiple European country by users choice.
4. **Vaccination Type Administered** – This chart to show percentage of total vaccine administrated by different manufacture vaccine type in Ireland



**Fig. 15.** Ireland Vaccination Dashboard – Year 2021

Limitations And Conclusion

Limitation

The major limitation is with the dataset of Age Group with peoples per hundred data set, the values are more than hundred for the groups above 60. After we raising (Ticket link in the foot note) *[[3]](#footnote-3) [[4]](#footnote-4)* the issue with “**Our World in Data (OWID)**” team, came to know that data is wrong with the “**European Centre for Disease Prevention and Control (ECDP)**” and formula they have used to calculation have limitation. OWID also given explanation that, data will be more than hundred because of multiple reasons they are vaccination of non-residents, old population estimate, etc. ECDP have collected the data from the officials, which is directly used in their system. ECDP website [[5]](#footnote-5) show the age group data till 60, because of this reason in the dashboard age group above 60 is showing the average value, which is more or less correct.

Dataset I have used only to meet the requirement of the KPI of this dashboard. There is range of dataset is available for all the countries. My dashboard limited to a single country Ireland.

Conclusion

This dashboard can give you an insight to the Ireland vaccination occurred in the year 2021. People per hundred values in the dashboard will give more information to the Govt. officials and the general public, the vaccination type chart will provide the percentage of types of vaccine used in Ireland. A comparison of Ireland and other European Union country will show the officials the position and percentage of vaccination given to the people.

References

1. https://community.tableau.com/s/question/0D54T00000C62HPSAZ/creating-donut-chart-in-tableau
2. https://www.tessellationtech.io/dashboard-design-essentials-dashboard-layout-and-formatting/
3. https://public.tableau.com/en-us/s/blog/2018/07/make-your-dashboards-even-more-beautiful-tableau-public-20182

1. https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations [↑](#footnote-ref-1)
2. https://population.un.org/wpp/ [↑](#footnote-ref-2)
3. https://github.com/owid/covid-19-data/issues/2173 [↑](#footnote-ref-3)
4. https://github.com/owid/covid-19-data/issues/2142 [↑](#footnote-ref-4)
5. https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html#age-group-tab [↑](#footnote-ref-5)