SQL Server Database, BI and Data Visualization Project With Agile Methods

Team-Five

SQL Server Authentication	
Server name	team-5.database.windows.net
Login	teamfive
Password	Group@5five
After connecting, use **pbistudygroup** email to sign in.	

Project Background

Assume that you work for the United Nations Food and Agriculture Organization (UN-FAO) as a data analyst. Global crop production data is collected from all countries in the world each year.   You're tasked with data cleaning and building visualizations that can be published on the organization's website in the future. You will use SQL Server as a DBMS and Power BI as a data visualization tool.

This project is broken down into two parts, which correspond to two releases:

1. Data import and cleaning
2. Data Visualization

Agile methodology will be used to implement this project. Accordingly, the project is divided into two Releases, four Sprints and seven User stories as shown in the table below.

[A short note about Agile Methodology]

- In Agile Methodology, a Project has multiple Releases, with each release delivering a business value to its customers.
- A Release contains one or more sprints.
- A Sprint, with a time boxed scope, implements one or more stories in that sprint.
- A Story, an end user driven business function may contain one or more use cases. Often in the business, a story is interchangeably used with a use case (two terms from different generations). A story tends to be driven more by the business user perspective whereas use cases used to represent the system perspective.

Please note that, while in this project the Releases, sprints and stories are presented with short time intervals and the tasks are relatively simple, in the real world enterprise you may encounter tasks take longer times. Releases can take 2-4 months, while a Sprint can be 2-4 weeks in most cases.

Release	Sprint	Story
Release 1 (R1)	Sprint 1 (SP1)	Story 1
		Story 2
	Sprint 2 (SP2)	Story 3
Release 2 (R2)	Sprint 3 (SP3)	Story 4
		Story 5
	Sprint 4 (SP4)	Story 6
		Story 7

Stories

Release 1: Sprint 1:

Story 1 (8 hrs – 1 story point): Create a database called **WorldFoodProduction**on Azure SQL Server and use the SSIS to import the data for CO2, Surface Temperature Anomaly, Maize Production and Maize Yield into the **WorldFoodProduction** database.

For this use the csv data sets included with this project. The CO2 Codebook is only for description purpose.

Story 2 (8 hrs – 1 story point): Use SSIS to import data for Rice Production, Rice Yield, Wheat Production and Wheat Yield

Release 1: Sprint 2:

Story 3 (40 hrs – 5 story points): Data Cleaning. After the data is imported to the SQL Server database, it should be cleaned to make it ready for the visualization.

Here are the requirements for data cleaning:

- First identify countries that exist in all the data sets. The final outcome from all the tables should contain those countries that exist in all data sets (the eight tables)
- Create a view of all the distinct countries that exist in all tables. Call it vw_Countries
- From each tables, delete all countries information that don't exist in the vw_countries
- Get the range of the years common to all the tables --- this needs more work
  - Few countries only have crop production and yield data between 2000 and 2018. And they should be removed – identify the countries and remove them from all list

CO2_Data:

- iso_code, country, year, co2 and population are the only columns needed.

Merge all data sets into one using country and year as a key

The final data should be ready for visualization and this completes Release 1.

Release 2: Data Visualization. Use Power BI to create the following visualizations

Sprint 3:

Story 4: (16 hrs – 2 story points): Create separate data visualization that shows line charts for each of the 8 data sets – CO2, Surface Temperature, Maize Prod and Yield, Rice Prod and Yield, Wheat Prod and Yield. Also Create parameterized line chart that uses country name as a parameter and shows all information about the country – use one parameter (country) and create a filter with that parameter.

Story 5: (16 hrs – 2 story points): Create world crop production, yield and population chart between the year ranges. Sum all countries products, take average of all countries yields – per crop type and take sum of population.

Sprint 4:

Story 6: (16hrs – 2 story point): Create forecasting model for production for future years – 2018 to 2025 for some select countries and the world.

Story 7: (8hrs – 1 story point): Create other charts out of the dataset that you believe can describe the data in a more elaborate and easy to comprehend the data easily.