Data Science CSCI 3320
Project: Fertility Rate 1960-2020

$Student 1:{ Mahmoud Z. Abu Sultan , ID: 120201068 };

$Student 2:{ Ahmad Al Nahal , ID: 120202122 };

Data mining

About the domain: Business Understanding.

About the domain...

- Fertility rate from 1960 to 2020, that is, the average fertility rate during each of those years for each country in the world.
- That is, the fertility rate is refers to the average number of children that mothers give birth to during their childbearing years.
- The fertility rate for each year is calculated based on the total number of births per woman during the year.
- From the general proportion of total births in one country ,the average fertility rates of women in the country are obtained.
- It is an important indicator that reflects the fertility level of the population , and thus the population growth can be predicted.

About the domain.

- All of this is assuming no net migration and an unchanged(fixed) mortality rate.
- All in line with prevailing fertility rates by age.
- Fertility rate is also an indicator of population growth.
- It all shows the extent of the impact of development on fertility rates, taking into account the number of deaths.

Data mining

About the data: Data Understanding.

About the data...

- The Dataset for: Fertility Rate of 186 Countries in 1960-2020.
- The fertility rate is usually expressed as the number of births per 1,000 women of childbearing age.
- The age of women is usually defined as ranging from 15 to 49 years.
- A fertility rate of 2.1 is considered the replacement level.
- Which means that each woman has on average enough children to replace herself and her partner.
- Fertility rate above replacement level may indicate population growth.
- While the fertility rate is below replacement level may indicate population decline.

Data Exploration to Data Preparation

Data Exploration

Data exploration...

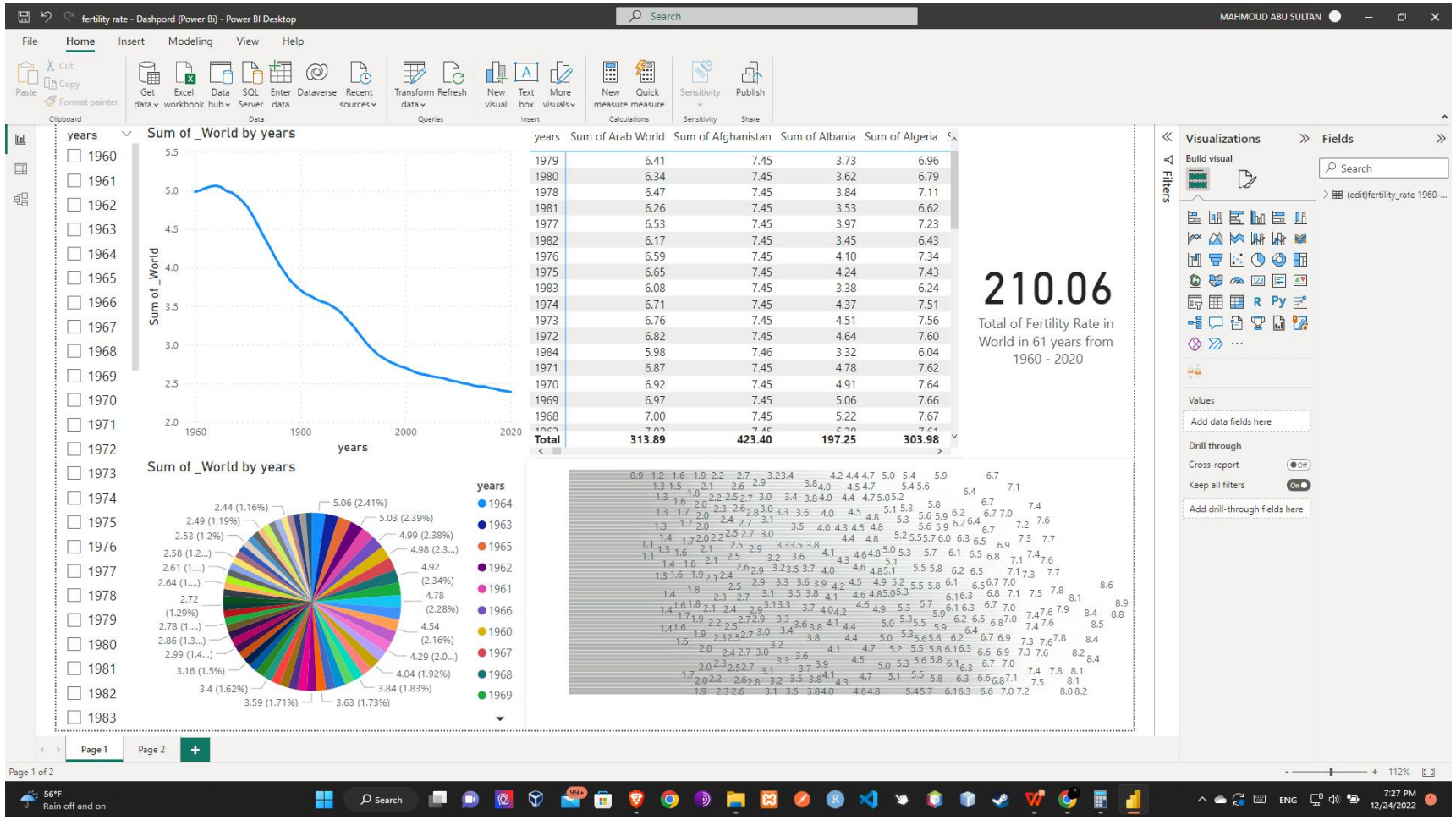
- Represented in 188 rows and 62 columns.
- where the rows(186) represent the countries and their fertility rates.
- Where there is a row(1) representing the fertility rate in percentage for the whole world during each year.
- And the columns(61) represent the years.
- There are no missing values. According to the current dataset from the source (kaggle.com), the ratio is (0%).

Data exploration...

- Useing descriptive statistics:
- The standard deviation, mean, mediand an provide a quick overview of the central tendency of the data and the diversity of the data.
  - The following values are calculated as a percentage of the total fertility rate in the worldfor 61 yers , 186 countres:
  - Mean = 3.44360
  - Median = 3.32
  - Range (2.39 - 4.98) = 2.59
  - Standard Deviation = 0.93
- Data visualization: -> in next page ...

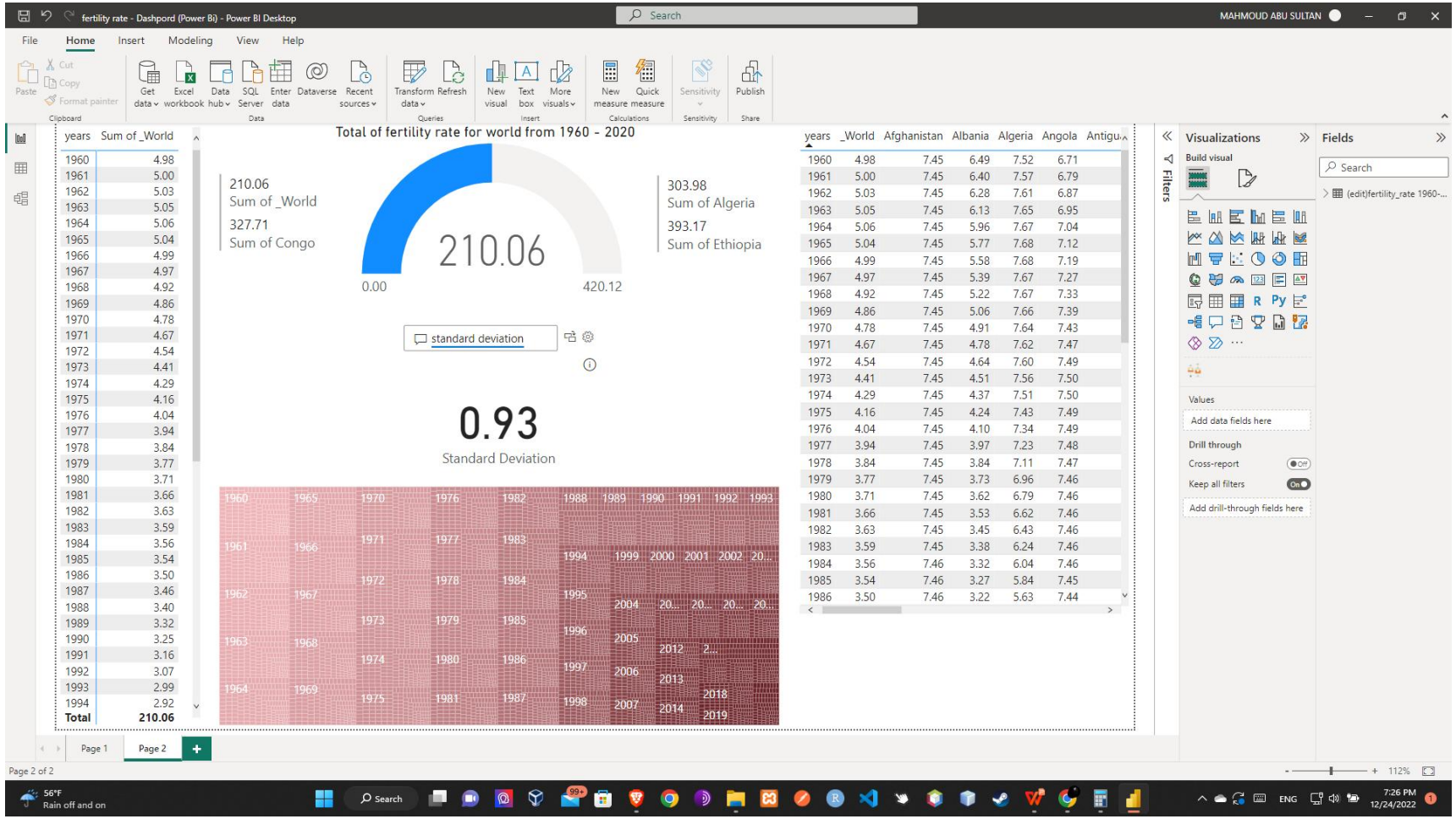
Data exploration…

Dashboard



Data exploration…

Dashboard



Data exploration.

Tell a story about what happened?

- The world has experienced a drastic decline in fertility rates from 1960 to 2020.
- Where the highest height reached was 5.06 in 1964.
- The Arab countries were the largest contributors to this increase ,as the fertility rate reached 7.05 in 1964.
- South Korea is the country with the lowest fertility rate from 1960 - 2020 estimated at 0.9 .
- Rwanda is the country with the highest fertility rate from 1960 - 2020 estimated at 8.9 .

Data Preparation

Data preparation.

- The data should be cleaned by deleting duplicate data.

  Duplicate in the Sudan data column.

- Replace empty cells with intermediate values or a value proportional to the string.
- Processing and repairing Missing data.part
- Since we are working on time series, we will split data based on time.
- We will use the data from the first half of the dataset for training ,

  And the data from the second half of the dataset for testing.

Part	# of records	000%
Train	30	50%
Test	30	50%

Business Analytics

Descriptive Analytics

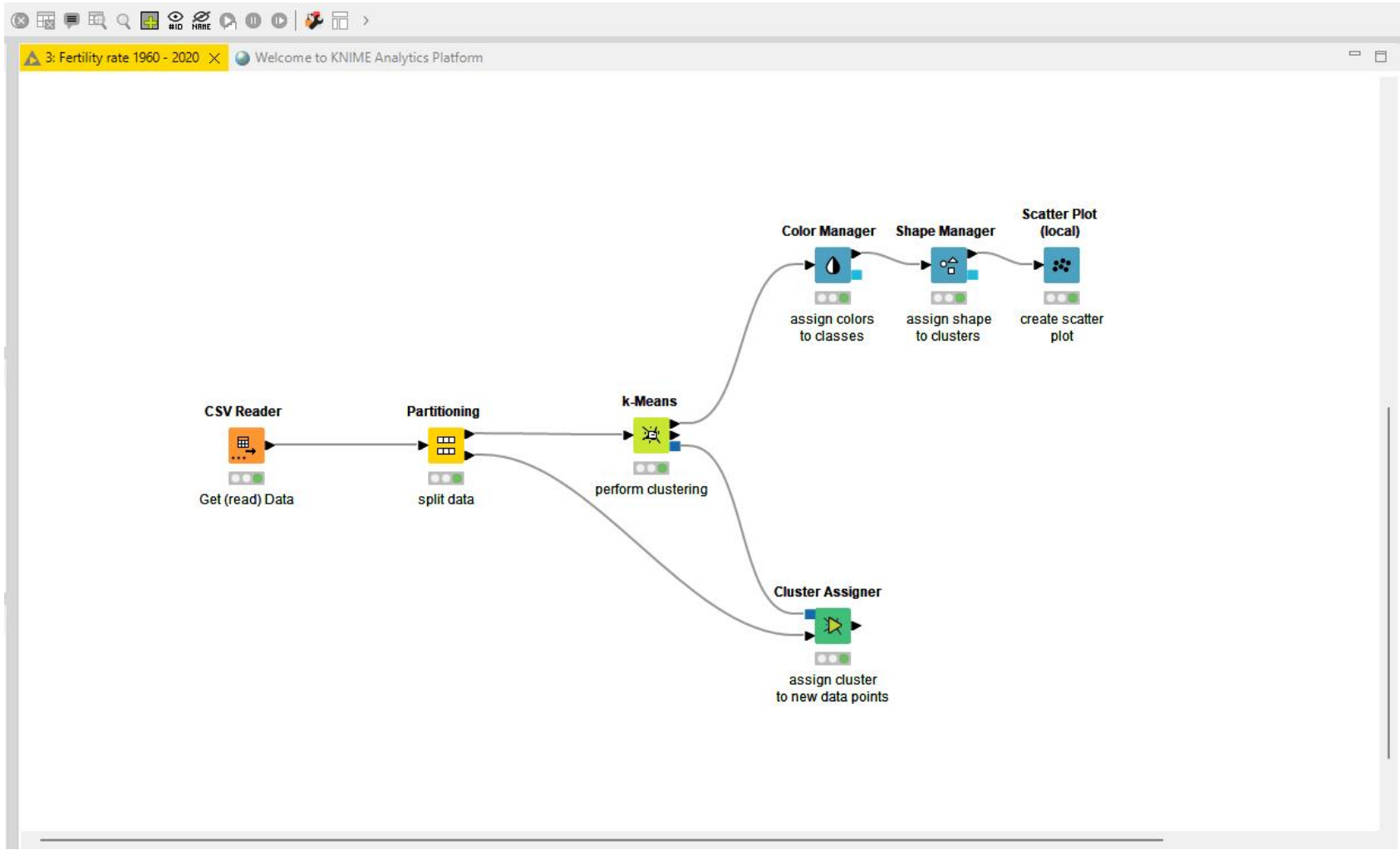
Descriptive analytics...

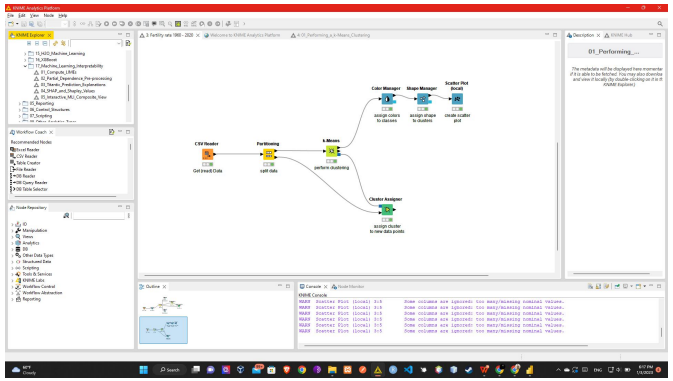
what happend? - what is happening?

- Fertility rates have decreased dramatically from 1960 to 2020.
- There is an increase in fertility rates at 1964 cents.
- The data is sufficient to obtain valid forecasts.

Descriptive analytics.

Use clustering to describe the data





Business Analytics

Predictive Analytics

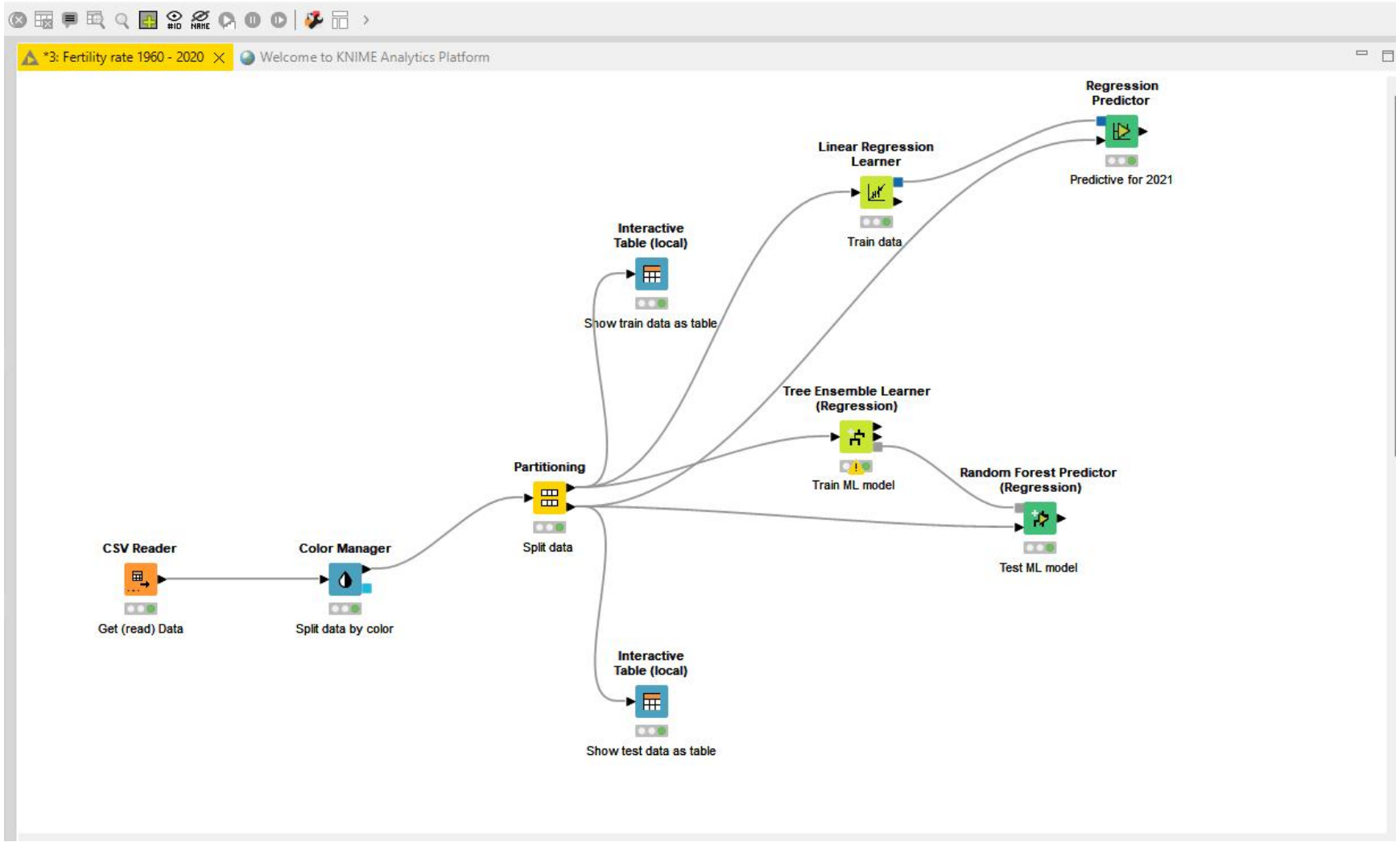
Predictive analytics...

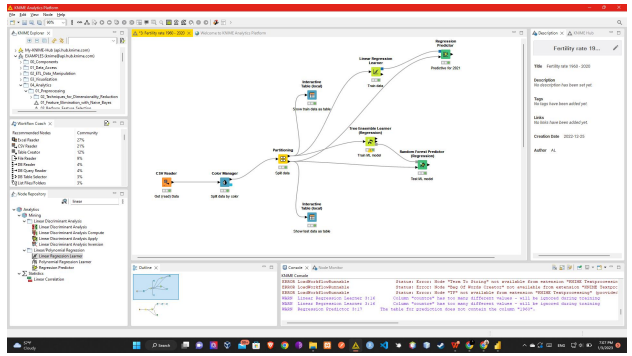
what will happen? - why will it happen?

- Fertility rates are predicted for 2021.
- The results were acceptable and in line with the time series.
- There will be an increase in fertility rates for the year.2021

Predictive analytics.

Use machine learning models to predict the future





Conclusions.

- Summarize your conclusions
- Only one slide


7M.T