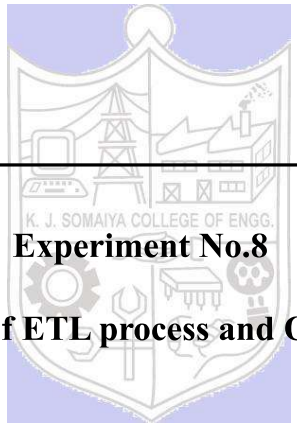



**Experiment No.8**

**Title: Execution of ETL process and OLAP operations**

**Batch: B-4** **Roll No.: 16010422234** **Name: Chandana Ramesh Galgali**

**Experiment No.: 8**

**Aim:** Execution of ETL process and OLAP operations

---

**Resources needed: Different RDBMS such as MySQL, Postgres and Excel, CSV, Rapidminer 5.3/ Latest version**

---

**Theory**

**Data Warehouse:**
An analytics-focused type of data management system called a data warehouse is intended to assist and allow business intelligence (BI) activities. Large amounts of historical data are frequently included in data warehouses, which are only designed to be used for queries and analysis. Application log files and transaction apps are only two examples of the many different sources from which the data in a data warehouse often comes.
Big data from various sources is centralised and combined in a data warehouse. Because of its analytical skills, businesses can get more out of their data and make better decisions. It gradually compiles a historical record that data scientists and business analysts can find quite useful. Because of these features, a data warehouse can be regarded as an organization's "single source of truth."

**ETL:**
Extract, Transform, Load (ETL) refers to a process in database usage and especially in data warehousing. Data extraction is where data is extracted from homogeneous or heterogeneous data sources; data transformation where the data is transformed for storing in the proper format or structure for the purposes of querying and analysis; data loading where the data is loaded into the final target database, more specifically, an operational data store, data mart, or data warehouse.

One may improve their chances of achieving better connection and scalability by employing a well-established ETL framework. A decent ETL tool must be able to interface with the several different relational databases and read the various file formats employed by a business. ETL solutions have started to move into Enterprise Application Integration, or even Enterprise Service Bus, systems that now encompass a lot more than simply the extraction, transformation, and loading of data. Converting CSV files into formats usable by relational databases is one frequent use case for ETL technologies. ETL solutions make it feasible for users to input csv-like data feeds/files and import it into a database with as little code as possible, facilitating a typical translation of millions of records. ESTL instruments

**RapidMiner:**
RapidMiner provides data mining and machine learning procedures including: data loading and transformation (Extract, transform, load (ETL)), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment. RapidMiner is written in the Java programming language. RapidMiner provides a GUI to design and execute analytical workflows. Those workflows are called “Processes” in RapidMiner and they consist of multiple “Operators”. Each operator performs a single task within the process, and the output of each operator forms the input of the next one. Alternatively, the engine can be

called from other programs or used as an API. Individual functions can be called from the command line. RapidMiner provides learning schemes, models and algorithms and can be extended using R and Python scripts.

**OLAP:**

In computing, online analytical processing, or OLAP is an approach to answering multi-dimensional analytical (MDA) queries. OLAP is part of the broader category of business intelligence, which also encompasses relational database report writing and data mining. Typical applications of OLAP include business reporting for sales, marketing, management reporting, business process management (BPM), budgeting and forecasting, financial reporting and similar areas, with new applications coming up, such as agriculture. The term OLAP was created as a slight modification of the traditional database term OLTP (Online Transaction Processing).

OLAP tools enable users to analyze multidimensional data interactively from multiple perspectives. OLAP consists of three basic analytical operations: consolidation (roll-up), drill-down, and slicing and dicing. Consolidation involves the aggregation of data that can be accumulated and computed in one or more dimensions. For example, all sales offices are rolled up to the sales department or sales division to anticipate sales trends. By contrast, the drill-down is a technique that allows users to navigate through the details. For instance, users can view the sales by individual products that make up a region's sales. Slicing and dicing is a feature whereby users can take out (slicing) a specific set of data of the OLAP cube and view (dicing) the slices from different viewpoints.

OLAP queries can be implemented by using analytical SQL functions.

Oracle has extensions to ANSI SQL to allow to quickly computing aggregations and rollups.
These new statements include:

rollup

cube

grouping

These simple SQL operators allow creating easy aggregations directly inside the SQL.

**Creating tabular aggregates with ROLLUP:**

ROLLUP enables an SQL statement to calculate multiple levels of subtotals across a specified group of dimensions. It also calculates a grand total. ROLLUP is a simple extension to the GROUP BY clause, so its syntax is extremely easy to use. Create cross-tabular reports with CUBE:

In multidimensional jargon, a "cube" is a cross-tabulated summary of detail rows. CUBE enables a SELECT statement to calculate subtotals for all possible combinations of a group of dimensions. It also calculates a grand total.

This is the set of information typically needed for all cross-tabular reports, so CUBE can calculate a cross-tabular report with a single select statement

---

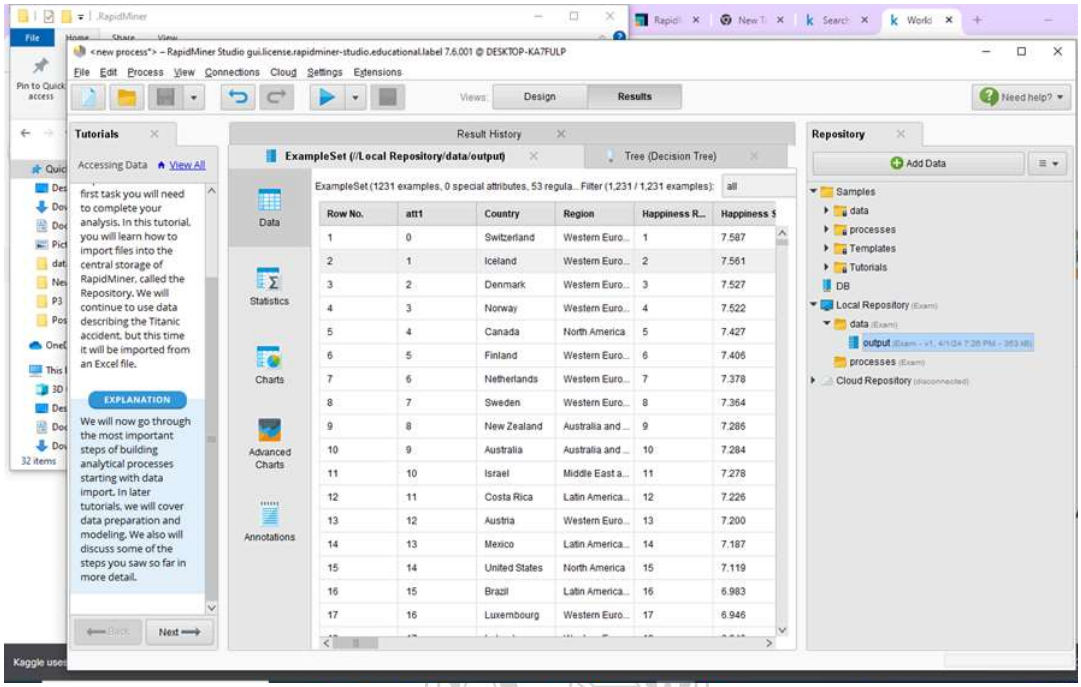
**Activities:**

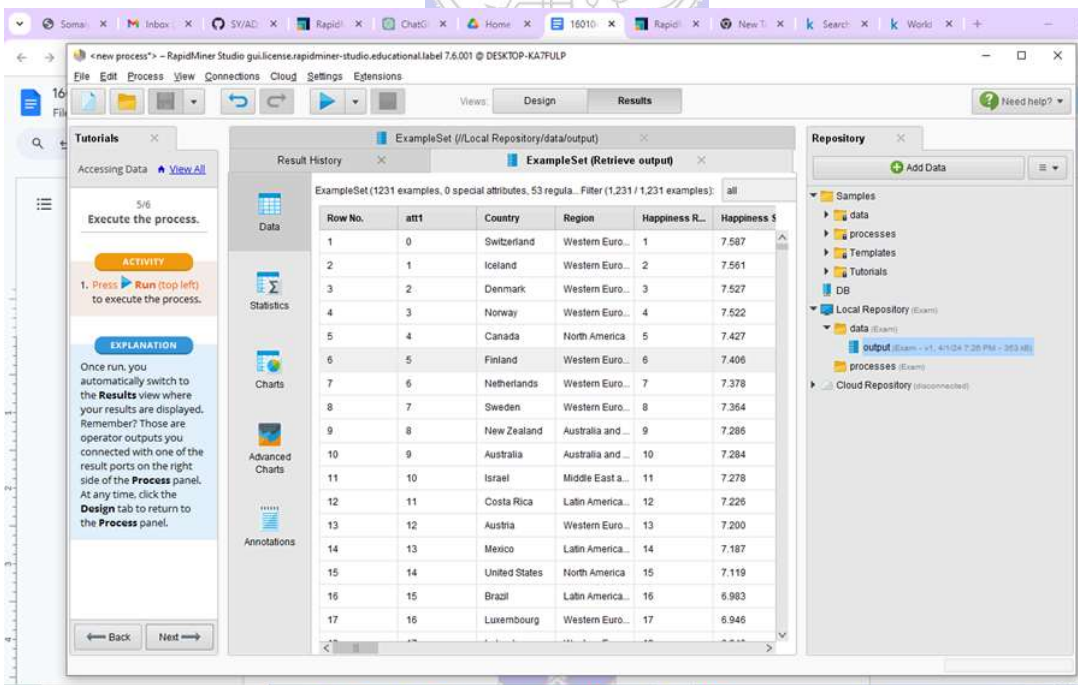
**For ETL and OLAP:**

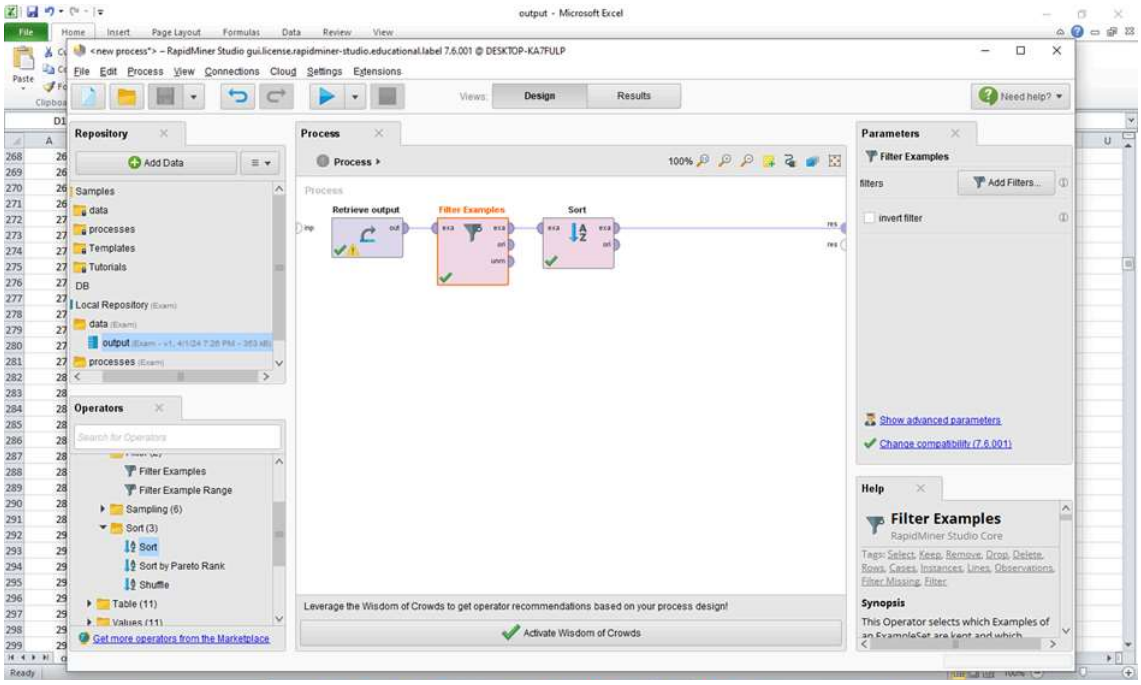
1. Install https://rapidminer.software.informer.com/download/#downloading
2. Go through the tutorial provided by RapidMiner

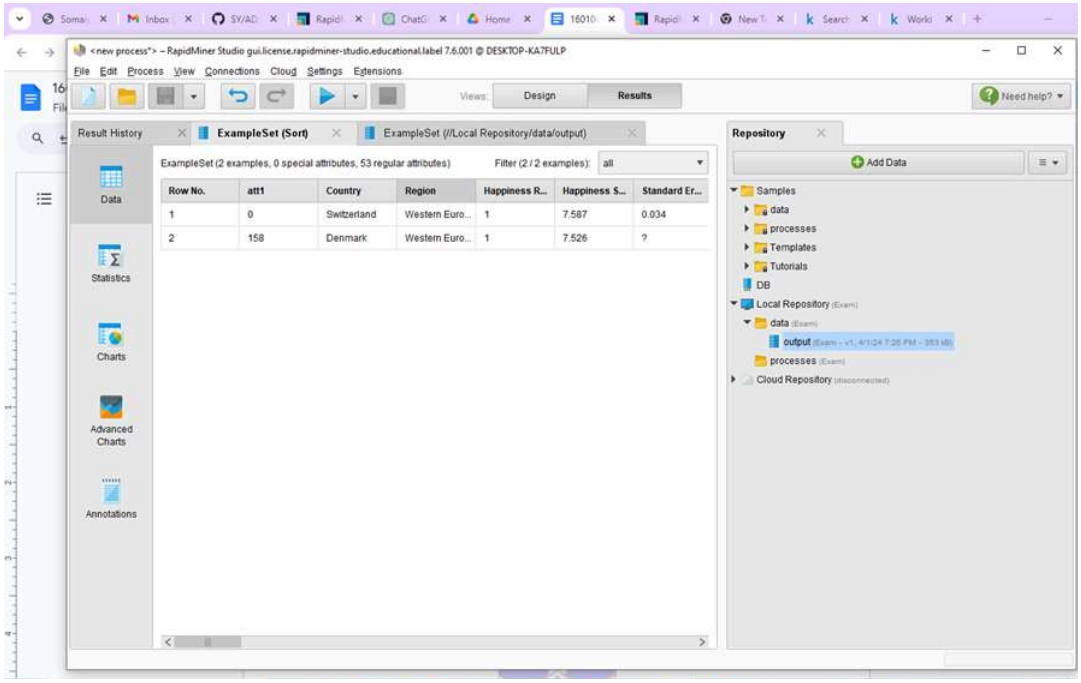
3. Extract data from 2 to 3 heterogeneous sources such as excel, MYSQL, Postgres etc.
4. Download any data set from *https://www.kaggle.com/datasets* or similar website
5. Apply five different transformations and filters to the data with specific requirement
6. Prepare a report for the activities 2 and 4 (ETL part) with steps and visualisations applied.
7. Create and save a clean dataset in a csv file.
8. Import the csv file from step7 in PostgreSQL database.
9. Apply rollup and cube operations to the same

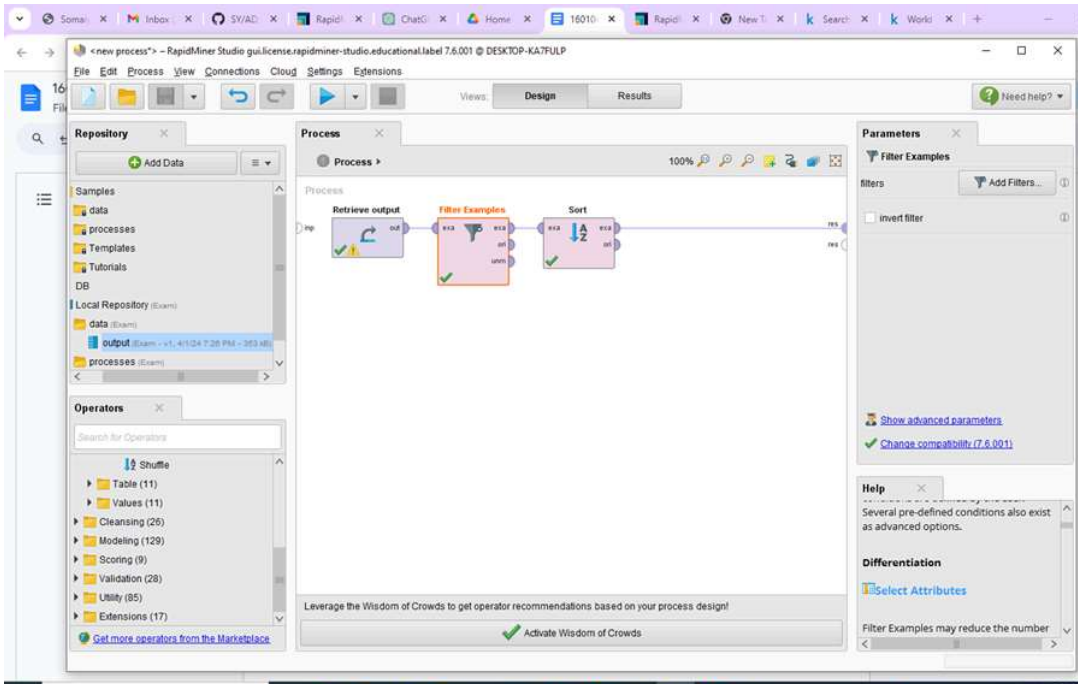
**ETL**

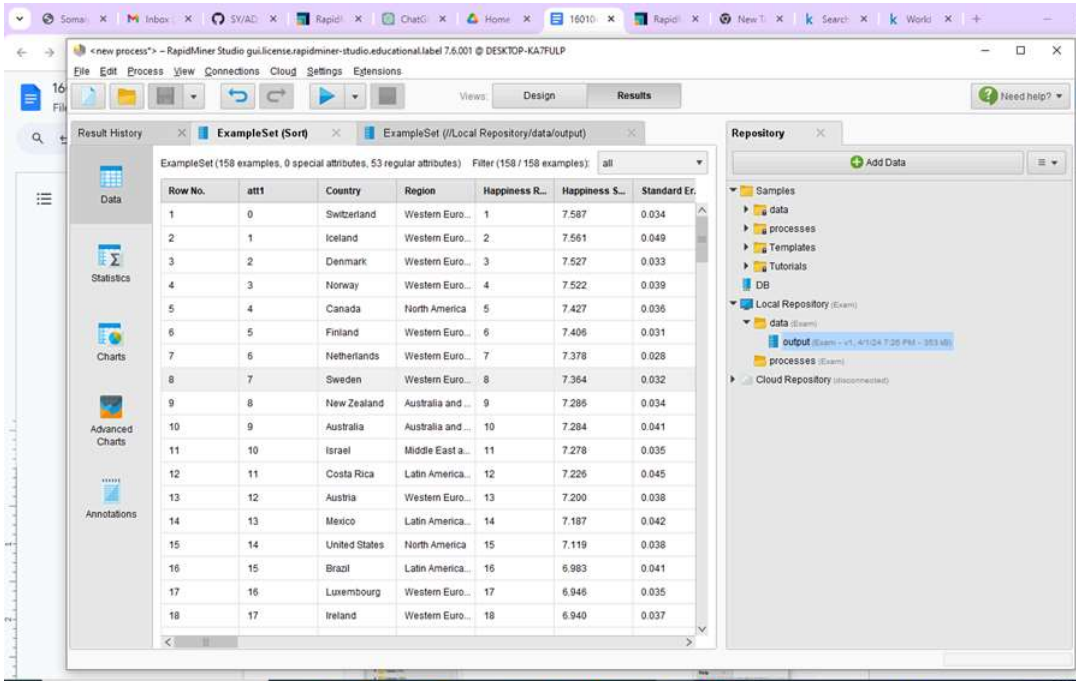


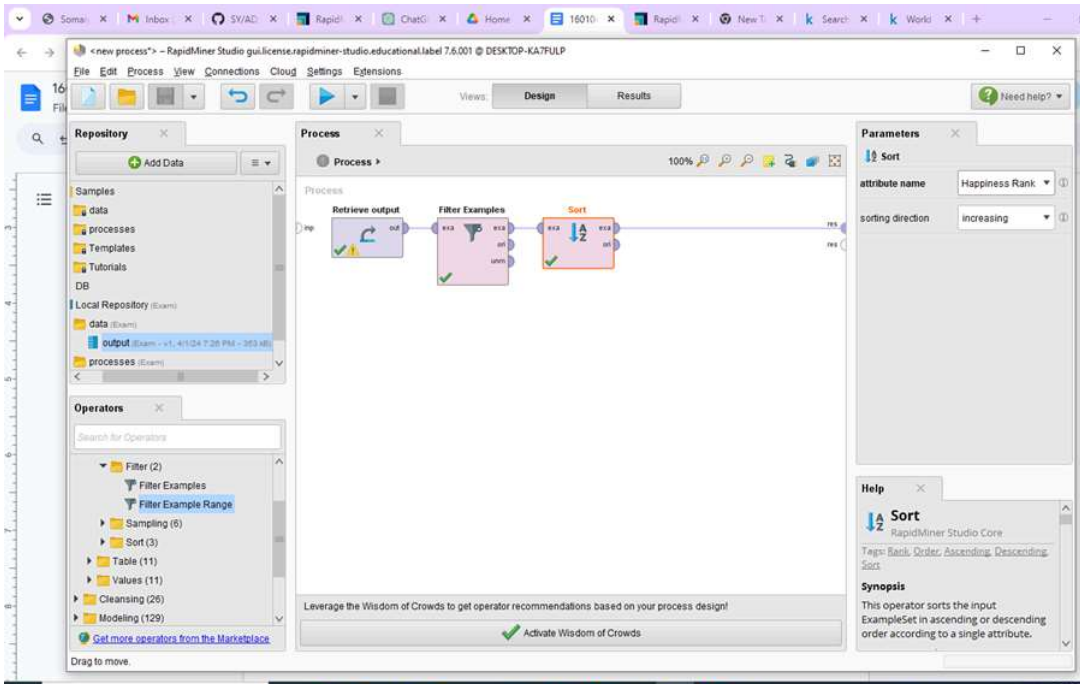



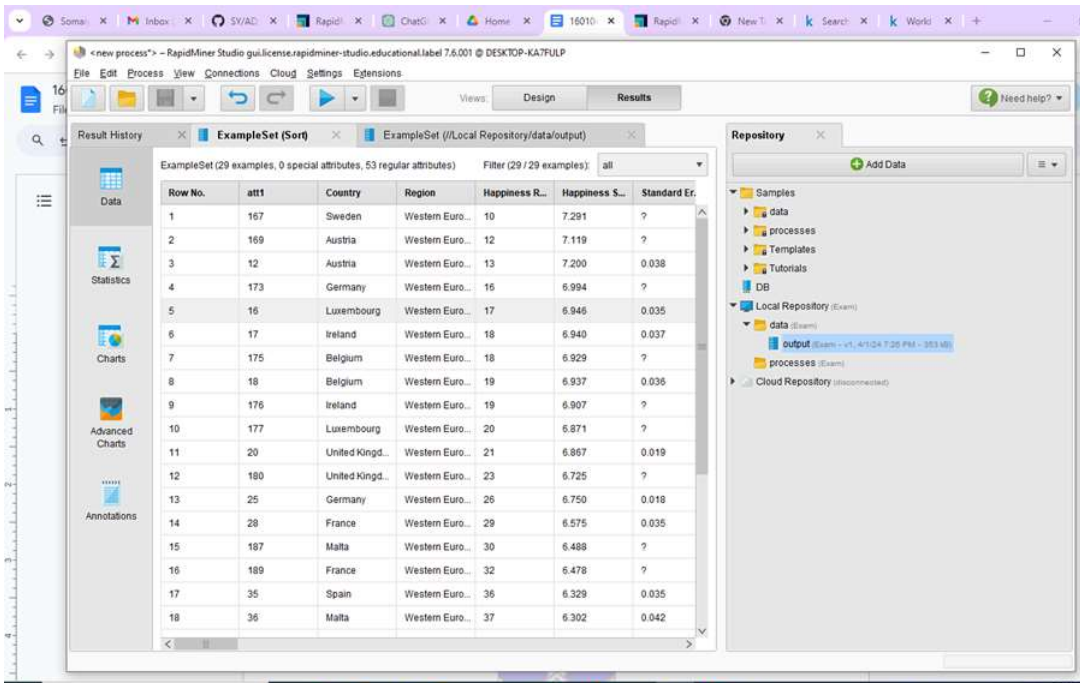

output - Microsoft Excel
<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Operators
Search for Operators
Filter Examples
Filter Example Range
Sampling (6)
Sort (3)
Sort
Sort by Pareto Rank
Shuffle
Table (11)
Values (11)
Get more operators from the Marketplace
Process
Process
100%
Retrieve output
Filter Examples
Sort
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Filter Examples
filters
Add Filters...
invert filter
Show advanced parameters
Change compatibility (7.6.001)
Help
Filter Examples
RapidMiner Studio Core
Tags: Select, Keep, Remove, Drop, Delete, Rows, Cases, Instances, Lines, Observations, Filter Missing, Filter
Synopsis
This Operator selects which Examples of
Ready

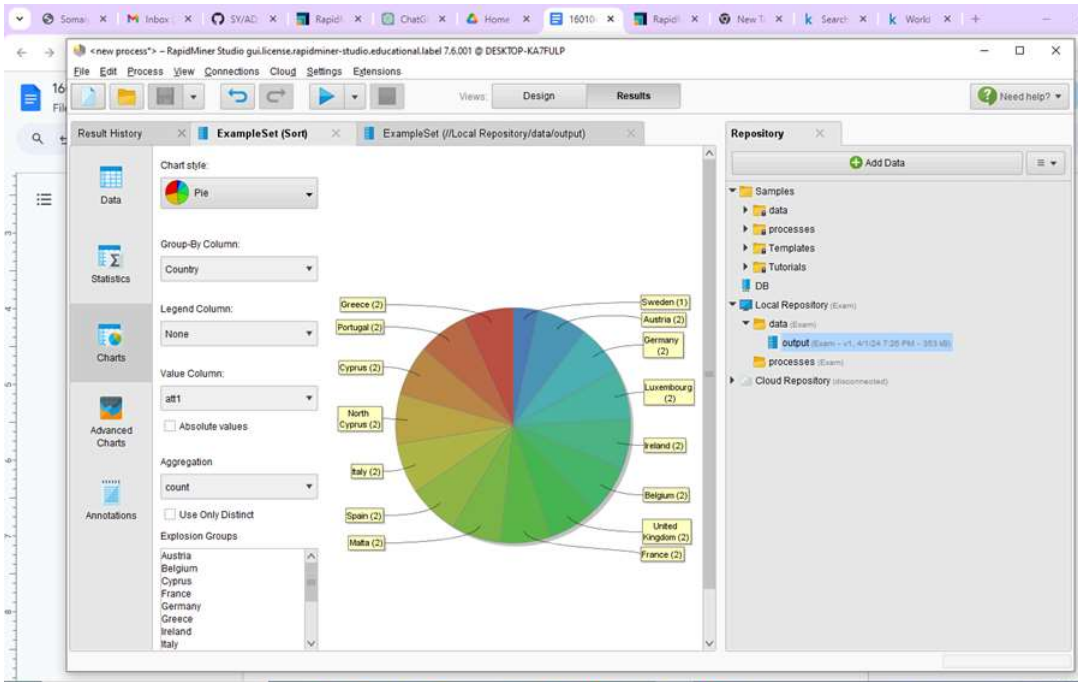

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
ExampleSet (2 examples, 0 special attributes, 53 regular attributes)
Filter (2 / 2 examples): all
Row No. att1 Country Region Happiness R... Happiness S... Standard Er...
1 0 Switzerland Western Euro... 1 7.587 0.034
2 158 Denmark Western Euro... 1 7.526 ?
Data
Statistics
Charts
Advanced Charts
Annotations
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Cloud Repository (disconnected)

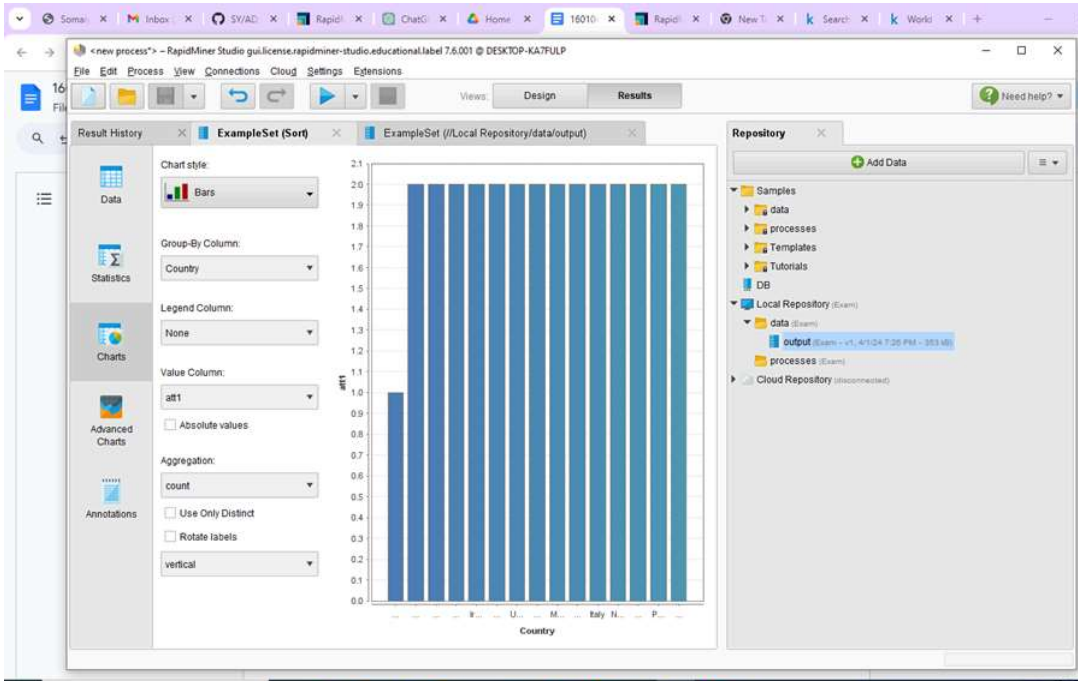

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Process
Process
100%
Retrieve output
Filter Examples
Sort
Operators
Search for Operators
Shuffle
Table (11)
Values (11)
Cleansing (26)
Modeling (129)
Scoring (9)
Validation (28)
Utility (85)
Extensions (17)
Get more operators from the Marketplace
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Filter Examples
filters
Add Filters...
invert filter
Show advanced parameters
Change compatibility (7.6.001)
Help
Several pre-defined conditions also exist as advanced options.
Differentiation
Select Attributes
Filter Examples may reduce the number

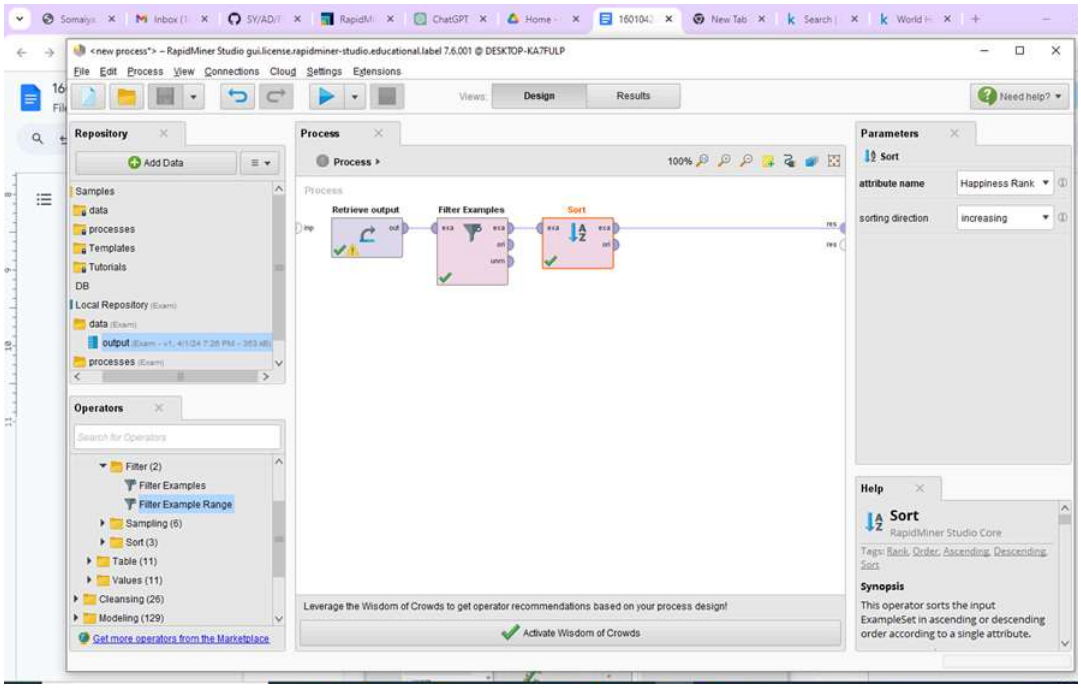

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
ExampleSet (158 examples, 0 special attributes, 53 regular attributes)
Filter (158 / 158 examples):
all
Data
Statistics
Charts
Advanced Charts
Annotations
Row No. | att1 | Country | Region | Happiness R... | Happiness S... | Standard Er.
1 | 0 | Switzerland | Western Euro... | 1 | 7.587 | 0.034
2 | 1 | Iceland | Western Euro... | 2 | 7.561 | 0.049
3 | 2 | Denmark | Western Euro... | 3 | 7.527 | 0.033
4 | 3 | Norway | Western Euro... | 4 | 7.522 | 0.039
5 | 4 | Canada | North America | 5 | 7.427 | 0.036
6 | 5 | Finland | Western Euro... | 6 | 7.406 | 0.031
7 | 6 | Netherlands | Western Euro... | 7 | 7.378 | 0.028
8 | 7 | Sweden | Western Euro... | 8 | 7.364 | 0.032
9 | 8 | New Zealand | Australia and ... | 9 | 7.286 | 0.034
10 | 9 | Australia | Australia and ... | 10 | 7.284 | 0.041
11 | 10 | Israel | Middle East a... | 11 | 7.278 | 0.035
12 | 11 | Costa Rica | Latin America... | 12 | 7.226 | 0.045
13 | 12 | Austria | Western Euro... | 13 | 7.200 | 0.038
14 | 13 | Mexico | Latin America... | 14 | 7.187 | 0.042
15 | 14 | United States | North America | 15 | 7.119 | 0.038
16 | 15 | Brazil | Latin America... | 16 | 6.983 | 0.041
17 | 16 | Luxembourg | Western Euro... | 17 | 6.946 | 0.035
18 | 17 | Ireland | Western Euro... | 18 | 6.940 | 0.037
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Cloud Repository (disconnected)

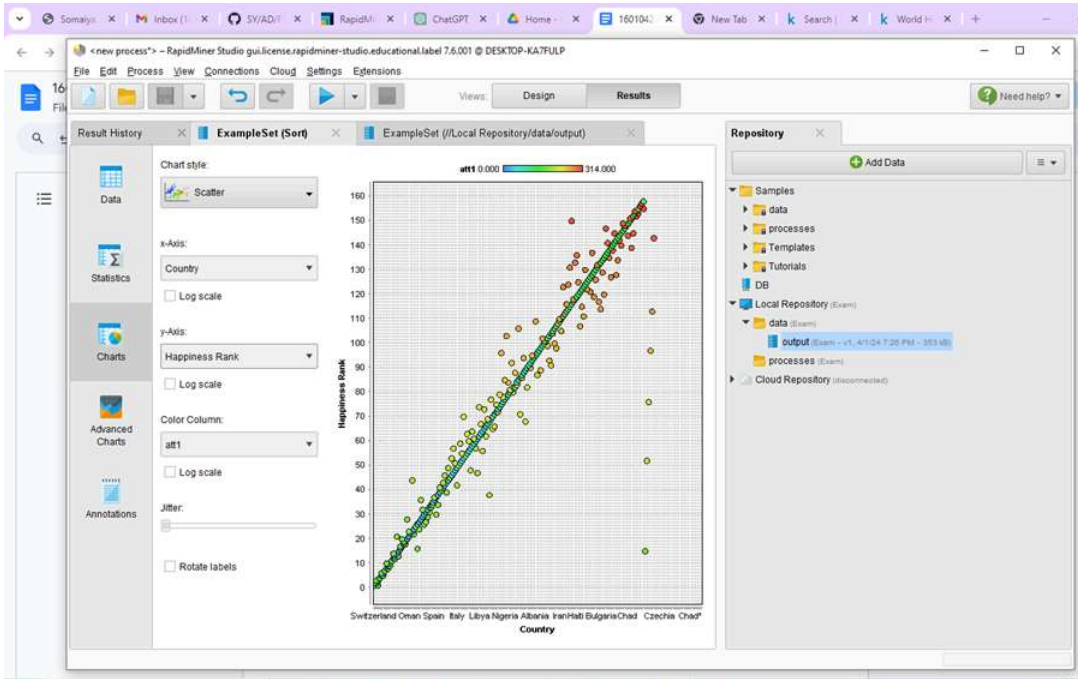

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Process
Process
100%
Retrieve output
Filter Examples
Sort
Parameters
Sort
attribute name
Happiness Rank
sorting direction
increasing
Operators
Search for Operators
Filter (2)
Filter Examples
Filter Example Range
Sampling (6)
Sort (3)
Table (11)
Values (11)
Cleansing (26)
Modeling (129)
Get more operators from the Marketplace
Help
Sort
RapidMiner Studio Core
Tags: Rank, Order, Ascending, Descending, Sort
Synopsis
This operator sorts the input ExampleSet in ascending or descending order according to a single attribute.
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Drag to move.

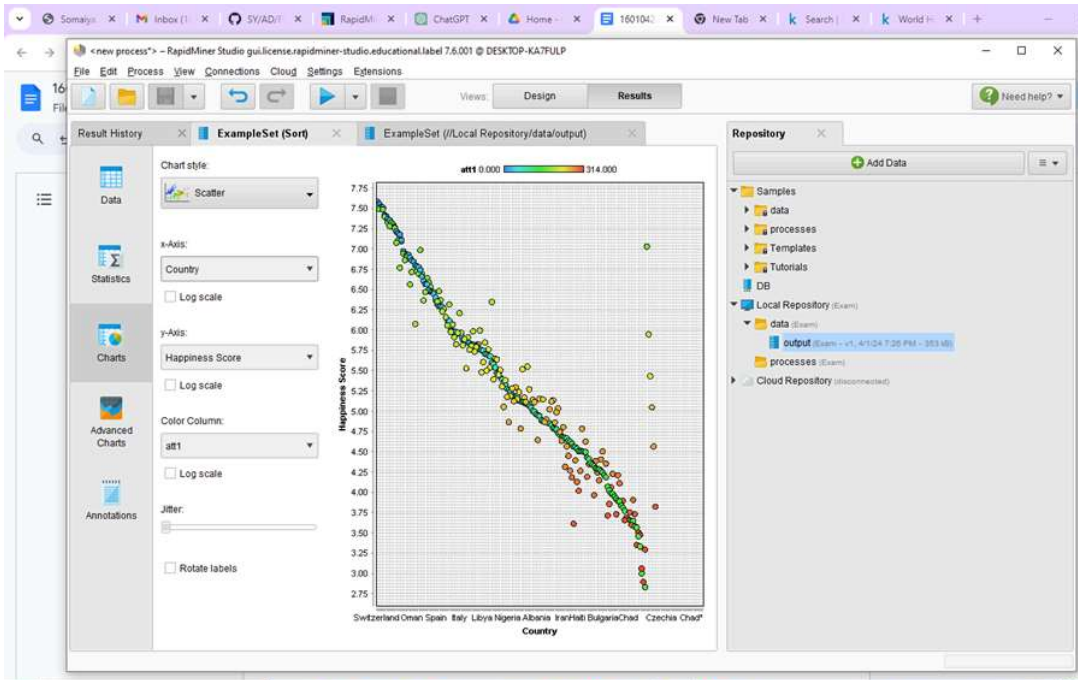

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
ExampleSet (29 examples, 0 special attributes, 53 regular attributes)
Filter (29 / 29 examples): all
Data
Statistics
Charts
Advanced Charts
Annotations
Row No. | att1 | Country | Region | Happiness R... | Happiness S... | Standard Er.
1 | 167 | Sweden | Western Euro... | 10 | 7.291 | ?
2 | 169 | Austria | Western Euro... | 12 | 7.119 | ?
3 | 12 | Austria | Western Euro... | 13 | 7.200 | 0.038
4 | 173 | Germany | Western Euro... | 16 | 6.994 | ?
5 | 16 | Luxembourg | Western Euro... | 17 | 6.946 | 0.035
6 | 17 | Ireland | Western Euro... | 18 | 6.940 | 0.037
7 | 175 | Belgium | Western Euro... | 18 | 6.929 | ?
8 | 18 | Belgium | Western Euro... | 19 | 6.937 | 0.036
9 | 176 | Ireland | Western Euro... | 19 | 6.907 | ?
10 | 177 | Luxembourg | Western Euro... | 20 | 6.871 | ?
11 | 20 | United Kingd... | Western Euro... | 21 | 6.867 | 0.019
12 | 180 | United Kingd... | Western Euro... | 23 | 6.725 | ?
13 | 25 | Germany | Western Euro... | 26 | 6.750 | 0.018
14 | 28 | France | Western Euro... | 29 | 6.575 | 0.035
15 | 187 | Malta | Western Euro... | 30 | 6.488 | ?
16 | 189 | France | Western Euro... | 32 | 6.478 | ?
17 | 35 | Spain | Western Euro... | 36 | 6.329 | 0.035
18 | 36 | Malta | Western Euro... | 37 | 6.302 | 0.042
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Cloud Repository (disconnected)

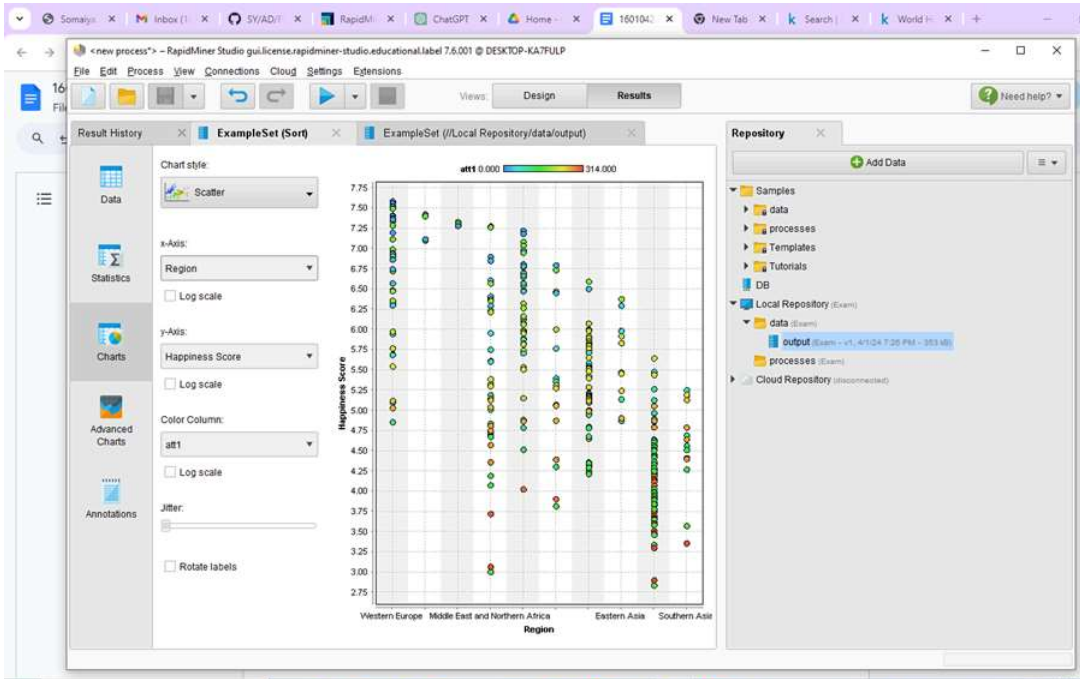

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
Data
Statistics
Charts
Advanced Charts
Annotations
Chart style:
Pie
Group-By Column:
Country
Legend Column:
None
Value Column:
att1
Absolute values
Aggregation
count
Use Only Distinct
Explosion Groups
Austria
Belgium
Cyprus
France
Germany
Greece
Ireland
Italy
Greece (2)
Portugal (2)
Cyprus (2)
North Cyprus (2)
Italy (2)
Spain (2)
Malta (2)
Sweden (1)
Austria (2)
Germany (2)
Luxembourg (2)
Ireland (2)
Belgium (2)
United Kingdom (2)
France (2)
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository
data
output
processes
Cloud Repository

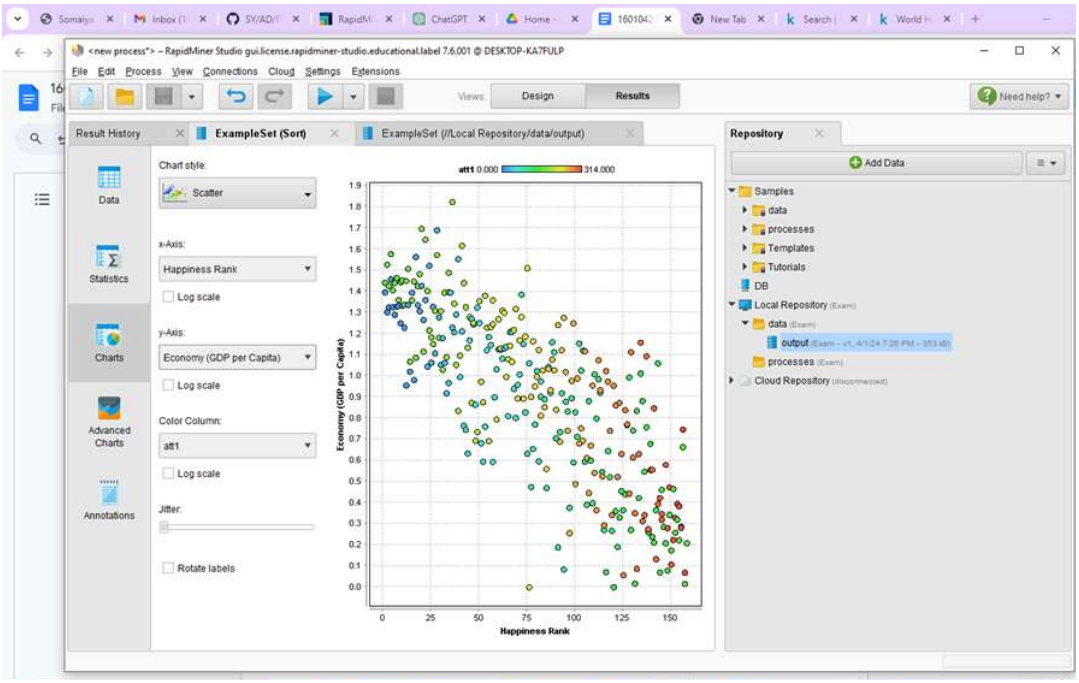

<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
Data
Statistics
Charts
Advanced Charts
Annotations
Chart style:
Bars
Group-By Column:
Country
Legend Column:
None
Value Column:
att1
Absolute values
Aggregation:
count
Use Only Distinct
Rotate labels
vertical
att1
Country
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository
data
output
processes
Cloud Repository


<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Operators
Search for Operators
Filter (2)
Filter Examples
Filter Example Range
Sampling (6)
Sort (3)
Table (11)
Values (11)
Cleansing (26)
Modeling (129)
Get more operators from the Marketplace
Process
Process »
100%
Retrieve output
Filter Examples
Sort
Leverage the Wisdom of Crowds to get operator recommendations based on your process design!
Activate Wisdom of Crowds
Parameters
Sort
attribute name
Happiness Rank
sorting direction
increasing
Help
Sort
RapidMiner Studio Core
Tags: Rank, Order, Ascending, Descending, Sort
Synopsis
This operator sorts the input ExampleSet in ascending or descending order according to a single attribute.


<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
Data
Statistics
Charts
Advanced Charts
Annotations
Chart style:
Scatter
x-Axis:
Country
Log scale
y-Axis:
Happiness Rank
Log scale
Color Column:
att1
Log scale
Jitter:
Rotate labels
att1 0.000 314.000
Happiness Rank
Country
Switzerland Oman Spain Italy Libya Nigeria Albania Iran Haiti Bulgaria Chad Czechia Chad*
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository (Exam)
data (Exam)
output (Exam - v1, 4/1/24 7:26 PM - 353 kB)
processes (Exam)
Cloud Repository (disconnected)


<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
Data
Statistics
Charts
Advanced Charts
Annotations
Chart style:
Scatter
x-Axis:
Country
Log scale
y-Axis:
Happiness Score
Color Column:
att1
Jitter:
Rotate labels
att1 0.000 314.000
Happiness Score
Country
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository
output
Cloud Repository


<new process*> – RapidMiner Studio gui.license.rapidminer-studio.educational.label 7.6.001 @ DESKTOP-KA7FULP
File Edit Process View Connections Cloud Settings Extensions
Views: Design Results
Need help?
Result History
ExampleSet (Sort)
ExampleSet (//Local Repository/data/output)
Data
Statistics
Charts
Advanced Charts
Annotations
Chart style:
Scatter
x-Axis:
Region
Log scale
y-Axis:
Happiness Score
Color Column:
att1
Jitter:
Rotate labels
att1 0.000 314.000
Happiness Score
Western Europe
Middle East and Northern Africa
Eastern Asia
Southern Asia
Region
Repository
Add Data
Samples
data
processes
Templates
Tutorials
DB
Local Repository
output
Cloud Repository




## OLAP

```
-- Step 1: Create Table
CREATE TABLE happiness_data (
    Country VARCHAR,
    Region VARCHAR,
    Happiness_Rank INTEGER,
    Happiness_Score NUMERIC,
    Standard_Error NUMERIC,
    Economy_GDP_per_Capita NUMERIC,
    Family NUMERIC,
    Health_Life_Expectancy NUMERIC,
    Freedom NUMERIC,
    Trust_Government_Corruption NUMERIC,
    Generosity NUMERIC,
    Dystopia_Residual NUMERIC,
    Year INTEGER,
    Lower_Confidence_Interval NUMERIC,
    Upper_Confidence_Interval NUMERIC,
    Happiness_Rank_2 INTEGER,
    Happiness_Score_2 NUMERIC,
    Whisker_High NUMERIC,
    Whisker_Low NUMERIC,
    Economy_GDP_per_Capita_2 NUMERIC,
    Health_Life_Expectancy_2 NUMERIC,
    Trust_Government_Corruption_2 NUMERIC,
    Dystopia_Residual_2 NUMERIC,
    Overall_Rank INTEGER,
    Country_or_Region VARCHAR,
    Score NUMERIC,
    GDP_per_capita NUMERIC,
    Social_support NUMERIC,
    Healthy_life_expectancy NUMERIC,
    Freedom_to_make_life_choices NUMERIC,
    Perceptions_of_corruption NUMERIC,
    Country_name VARCHAR,
    Regional_indicator VARCHAR,
    Ladder_score NUMERIC,
    Standard_error_of_ladder_score NUMERIC,
    Upper_whisker NUMERIC,
    Lower_whisker NUMERIC,
    Logged_GDP_per_capita NUMERIC,
    Ladder_score_in_Dystopia NUMERIC,
```

```
    Explained_by_Log_GDP_per_capita NUMERIC,
    Explained_by_Social_support NUMERIC,
    Explained_by_Healthy_life_expectancy NUMERIC,
    Explained_by_Freedom_to_make_life_choices NUMERIC,
    Explained_by_Generosity NUMERIC,
    Explained_by_Perceptions_of_corruption NUMERIC,
    Dystopia_residual NUMERIC,
    Rank INTEGER,
    Happiness_score_3 NUMERIC,
    Whisker_high_2 NUMERIC,
    Whisker_low_2 NUMERIC,
    Dystopia_1_83_residual NUMERIC,
    Explained_by_GDP_per_capita NUMERIC
);

-- Step 2: Import CSV Data
COPY happiness_data FROM '"C:\Users\Exam\Downloads\output.csv"' DELIMITER
',' CSV HEADER;

-- Step 3: Apply Rollup Operation
-- Rollup Operation on Happiness Rank and Region
SELECT Country, Region, SUM(Happiness_Score) AS Total_Happiness_Score
FROM happiness_data
GROUP BY ROLLUP(Country, Region);

-- Step 4: Apply Cube Operation
-- Cube Operation on Happiness Rank, Country, and Year
SELECT Happiness_Rank, Country, Year, SUM(Happiness_Score) AS
Total_Happiness_Score
FROM happiness_data
GROUP BY CUBE(Happiness_Rank, Country, Year);
```

| Country | Region | Total_Happiness_Score |
|---|---|---|
| Afghanistan | Central Asia | 25.6 |
| Afghanistan | NULL | 25.6 |
| Albania | Eastern Europe | 38.9 |
| Albania | NULL | 38.9 |
| NULL | NULL | 64.5 |

| Happiness_Rank | Country | Year | Total_Happiness_Score |
|---|---|---|---|
| 1 | Afghanistan | 2019 | 5.653 |
| 1 | Afghanistan | NULL | 5.653 |
| 1 | NULL | 2019 | 5.653 |
| 1 | NULL | NULL | 5.653 |
| 2 | Afghanistan | 2018 | 5.791 |
| 2 | Afghanistan | NULL | 5.791 |
| 2 | NULL | 2018 | 5.791 |
| 2 | NULL | NULL | 5.791 |
| ... | ... | ... | ... |
| NULL | NULL | NULL | 64.5 |

**Questions:**

**1. Elaborate on the operations applied and results generated to your dataset.**

**Ans:** The operations applied in the ETL process aimed at preparing the data for analysis by ensuring its quality, consistency, and relevance. Each transformation addressed specific aspects of data preprocessing, such as handling missing values, standardizing formats, and creating new features.

For instance, normalization helped in standardizing numerical attributes, ensuring that they contribute proportionally to the analysis without biases due to scale. Missing value handling techniques ensured that no information was lost during analysis, maintaining the integrity of the dataset.

The visualizations included in the report provided insights into the distribution of data, relationships between variables, and patterns within the dataset. These visualizations aided in understanding the effects of transformations and identifying potential trends or outliers.

**2. Explain if Drill-down, Drill-across can be applied in relational databases, Justify with a query implementation.**

**Ans:** While OLAP operations like drill-down and drill-across are traditionally associated with multidimensional databases, similar functionalities can be achieved in relational databases through SQL queries.

For example, consider the following SQL query to implement drill-down:

```
SELECT Category, Product, SUM(Sales)
FROM SalesData
GROUP BY ROLLUP(Category, Product);
```

This query calculates subtotals for each category and product combination, as well as total sales for each category. By removing the Product column from the GROUP BY clause, the query can drill down to obtain subtotals at the product level within each category.

Similarly, drill-across operations can be achieved by joining tables from different dimensions or hierarchies in relational databases, allowing analysts to explore relationships across different dimensions.

---

**Results:**

**Report for ETL**

Steps and Visualizations Applied:

1. Data Extraction: Data was extracted from heterogeneous sources including Excel, MySQL, and PostgreSQL. A dataset was downloaded from Kaggle for further processing.

2. Data Transformation and Filters:
   - Normalization: Ensured that all data attributes fall within a similar range to avoid biases in analysis.
   - Missing Value Handling: Implemented strategies like imputation or deletion to handle missing values.
   - Data Aggregation: Aggregated data to higher levels to provide summaries for analysis.
   - Data Cleaning: Removed duplicates, corrected inconsistencies, and standardized formats for uniformity.
   - Feature Engineering: Created new features to enhance analysis, such as calculating ratios or adding derived attributes.

3. Report Preparation: A comprehensive report was prepared detailing the steps taken in the ETL process. Visualizations such as histograms, scatter plots, and bar charts were included to illustrate transformations and patterns in the data.

4. Clean Dataset Creation: After applying transformations and filters, a clean dataset was generated and saved in CSV format. This dataset serves as a reliable input for further analysis.

5. Import to PostgreSQL: The cleaned dataset was imported into a PostgreSQL database for storage and efficient querying.

6. OLAP Operations:
    - Rollup: Utilized the ROLLUP SQL extension to calculate multiple levels of subtotals across specified dimensions.
    - Cube: Employed the CUBE extension to calculate subtotals for all possible combinations of dimensions, facilitating cross-tabular reports.

---

**Outcomes: Apply ETL processing and Online Analytical Processing on the warehouse data.**

---

**Conclusion: (Conclusion to be based on the outcomes achieved)**
The ETL process successfully transformed raw data from heterogeneous sources into a clean and structured dataset suitable for analysis. By applying various transformations and filters, data quality was enhanced, ensuring the reliability of insights derived from the dataset. OLAP operations further facilitated interactive analysis and reporting, enabling users to gain valuable insights from multidimensional data.



---

**Grade: AA / AB / BB / BC / CC / CD /DD**

**Signature of faculty in-charge with date**

---

**References:**


- https://www.oracle.com/in/database/what-is-a-data-warehouse
- Paulraj Ponniah, “Data Warehousing: Fundamentals for IT Professionals”, Wiley India