

# TUTORIAL 1

# ORANGE, RAPIDMINER, AND

# DATA PREPROCESSING

---

Ari, CS528, 2026, NS

# Con ten ts

---

- [Orange](#)
- [RapidMiner](#)
- [Data Processing Examples](#)



Orange



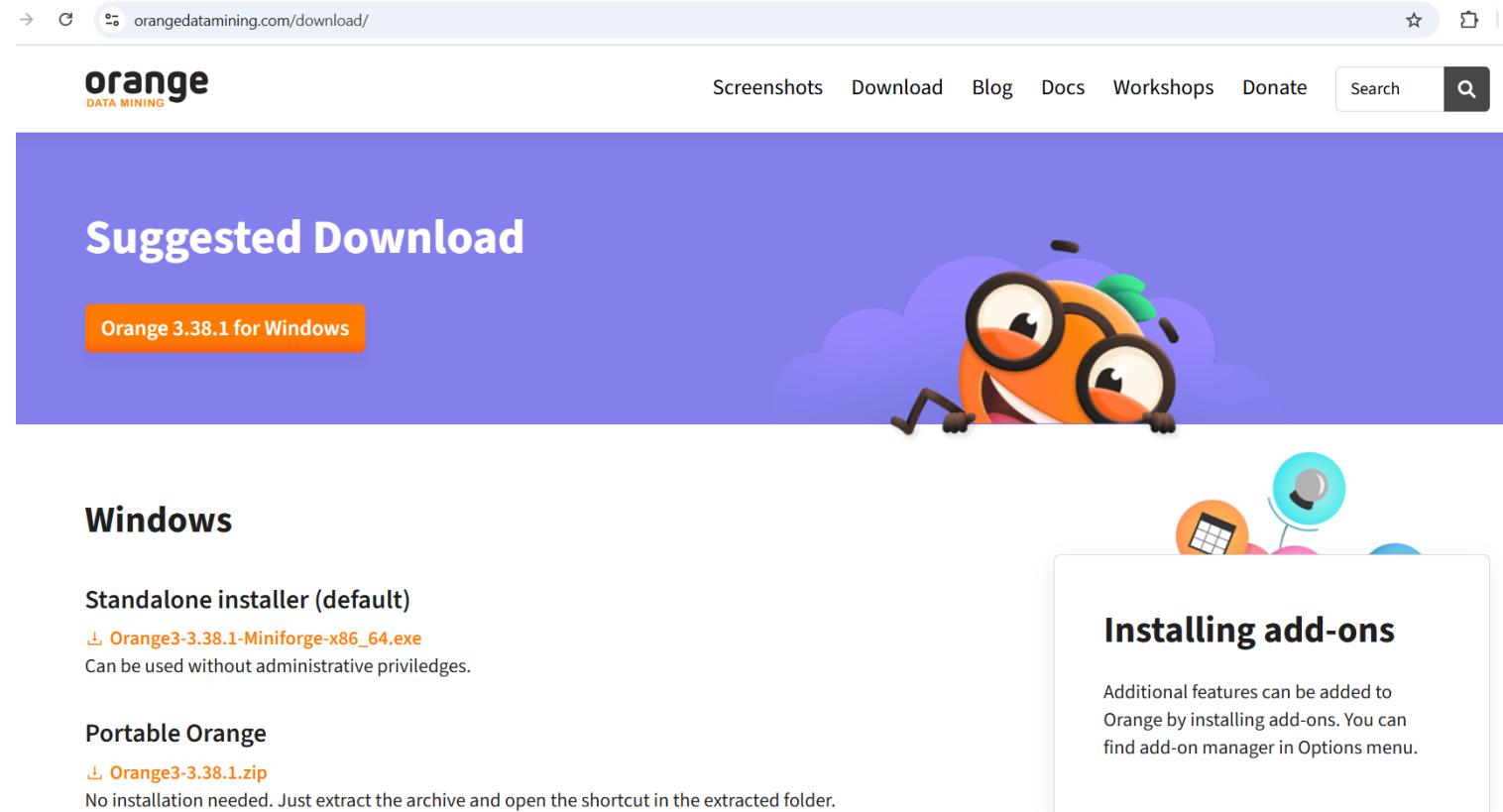
# Orange

- Orange is a powerful platform to perform data analysis and visualization, see data flow and become more productive. It provides a clean, open-source platform and the possibility to add further functionality for all fields of science.

The screenshot shows the homepage of the Orange Data Mining website. At the top left is the "orange DATA MINING" logo. To its right is a navigation bar with links: Screenshots, Download, Blog, Docs, Workshops, and Donate. Next is a search bar with a magnifying glass icon. The main content area has a purple background. On the left, the text "Data Mining Fruitful and Fun" is displayed in large white font, with "Open source machine learning and data visualization." below it. A call-to-action button at the bottom left says "Download Orange 3.38.1". On the right side, there is a cartoon illustration of an orange character wearing glasses, surrounded by various data visualization icons like a lightbulb, a grid, and a network graph.

# Or an ge, Down loa d a n d In sta ll

- <https://orangedatamining.com/download/>



The screenshot shows the official website for Orange Data Mining. At the top, there's a navigation bar with links for Screenshots, Download, Blog, Docs, Workshops, and Donate, along with a search bar. Below the header, a large purple banner features a cartoon orange character with glasses and a smiling face, peeking out from behind clouds. The banner has the text "Suggested Download" and a button labeled "Orange 3.38.1 for Windows". The main content area is divided into sections for "Windows" and "Portable Orange". Under "Windows", there's a link to a standalone installer ("Orange3-3.38.1-Miniforge-x86\_64.exe") which is described as being used without administrative privileges. Under "Portable Orange", there's a link to "Orange3-3.38.1.zip" which is described as requiring no installation, just extraction and opening the shortcut. To the right, there's a callout box titled "Installing add-ons" explaining how additional features can be added via the Options menu.

orange  
DATA MINING

Screenshots Download Blog Docs Workshops Donate Search

## Suggested Download

Orange 3.38.1 for Windows

### Windows

**Standalone installer (default)**  
[Orange3-3.38.1-Miniforge-x86\\_64.exe](#)  
Can be used without administrative priviledges.

**Portable Orange**  
[Orange3-3.38.1.zip](#)  
No installation needed. Just extract the archive and open the shortcut in the extracted folder.

### Installing add-ons

Additional features can be added to Orange by installing add-ons. You can find add-on manager in Options menu.

# Rapidm in er

---



## Depth for Data Scientists, Simplified for Everyone Else

Join the 40,000+ global organizations in every industry who use the RapidMiner data science platform to drive revenue, reduce costs, and avoid risk.

- RapidMiner provides data mining and machine learning procedures including: data loading and transformation (ETL), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation, and deployment.
- RapidMiner provides a GUI to design and execute analytical workflows.



# RAPIDMINER



## The Hard Way

- › Slow to collaborate & transform
- › Acquire highly specialized expertise
- › Choose complexity or oversimplification
- › Opaque, fragmented processes
- › Commit to a specific ecosystem

## The RapidMiner Way

- › Transformational business impact
- › Upskill your organization
- › Depth for data scientists, simplified for everyone else
- › Easy to trust, tune & explain
- › Future-proof innovation, portability & extensibility



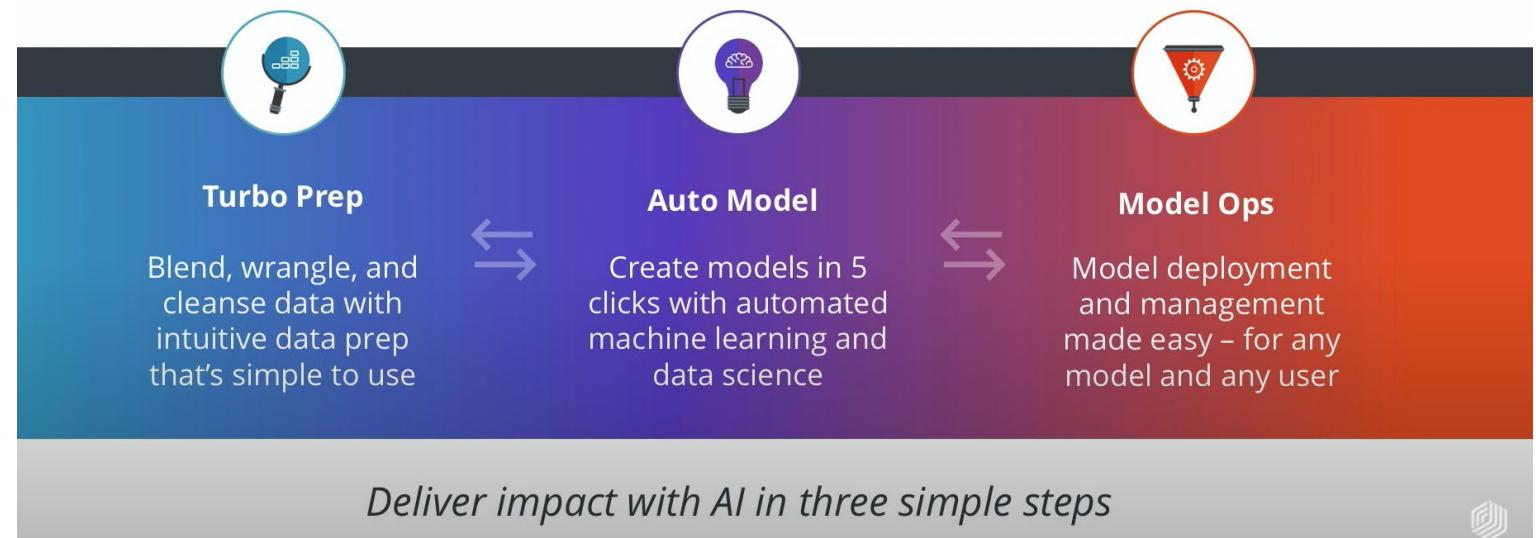
# RAPIDMINER



[WWW.RAPIDMINER.COM](http://WWW.RAPIDMINER.COM),

[WWW.YOUTUBE.COM/RAPIDMINER](http://WWW.YOUTUBE.COM/RAPIDMINER)

## RapidMiner Offers a Complete Path to Fully Automated Data Science



# RAPIDMINER

•  HEX

## FORRESTER®

RapidMiner is a Leader in The Forrester Wave: Multimodal Predictive Analytics & Machine Learning Solutions, Q3 2020

## Gartner

RapidMiner is a Visionary in the 2021 Gartner Magic Quadrant for Data Science and Machine Learning Platforms



RapidMiner is a June 2020 Gartner Peer Insights Customers' Choice for Data Science and Machine Learning Platforms for the third time in a row

## G2 CROWD

RapidMiner is the Highest Rated, Easiest to Use Data Science and Machine Learning Platform and was named a Leader in G2's Spring 2021 Report.



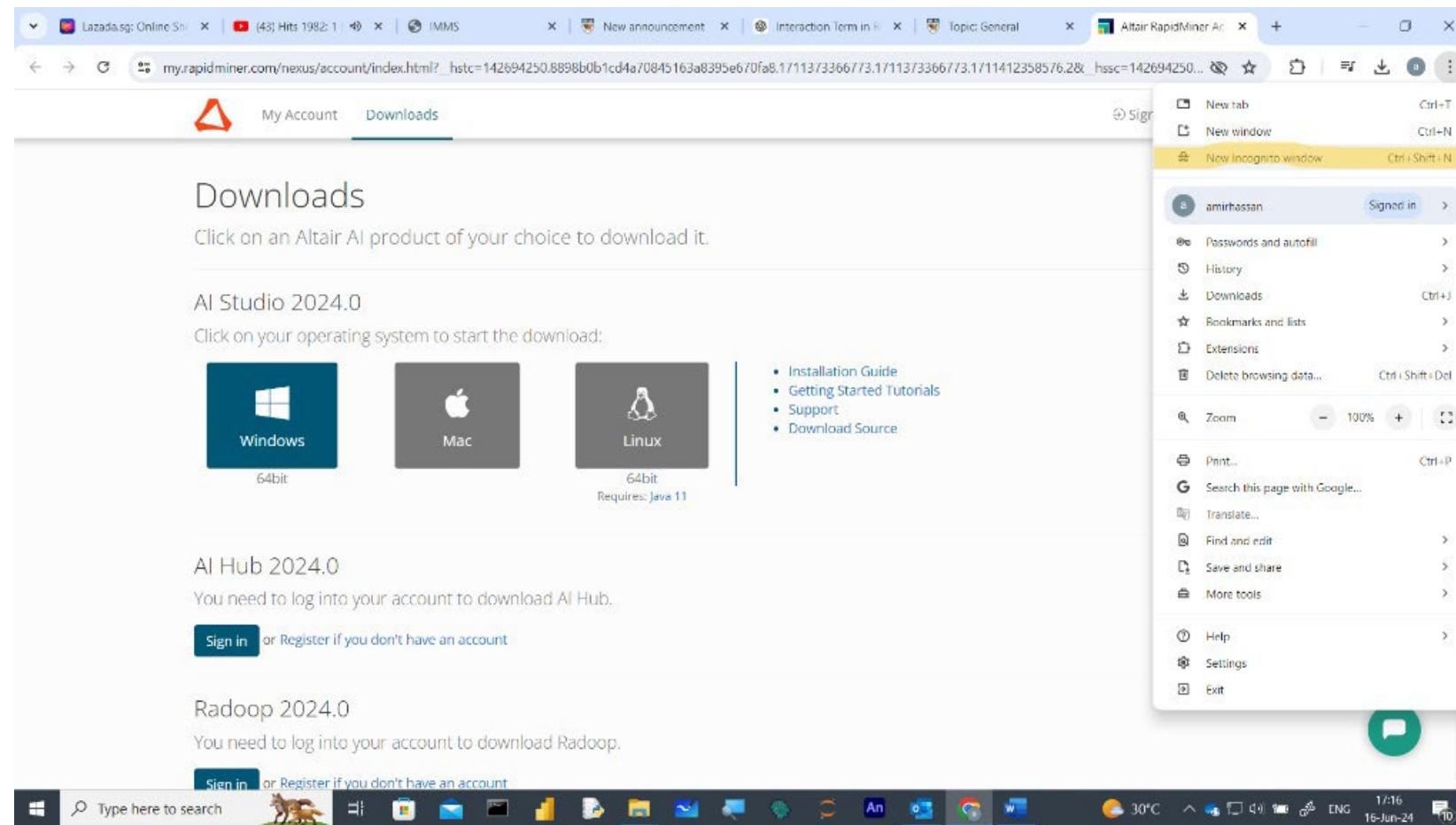
Rapidminer

# Download and Install

---

- To download Rapidminer:
- [https://my.rapidminer.com/nexus/account/index.html?\\_hstc=142694250.8898b0b1cd4a70845163a8395e670fa8.1711373366773.1711373366773.1711412358576.2&\\_hssc=142694250.1.1711412358576&\\_hsfp=676704640](https://my.rapidminer.com/nexus/account/index.html?_hstc=142694250.8898b0b1cd4a70845163a8395e670fa8.1711373366773.1711373366773.1711412358576.2&_hssc=142694250.1.1711412358576&_hsfp=676704640)
- You need to have a Rapidminer account.
- Try this website to create a Rapidminer account instead of the main webpage:
- [https://my.rapidminer.com/nexus/account/index.html?\\_hstc=142694250.3188faa95341077fb0a799ae44e201f4.1714093534218.1714093534218.1714096391239.2&\\_hssc=142694250.4.1714096391239&\\_hsfp=2414989234#signup](https://my.rapidminer.com/nexus/account/index.html?_hstc=142694250.3188faa95341077fb0a799ae44e201f4.1714093534218.1714093534218.1714096391239.2&_hssc=142694250.4.1714096391239&_hsfp=2414989234#signup)
- Tick "For Educational Purpose"
- Using incognito sometimes is better if your browser is Chrome.

# Download and Install



# Download and Install

The screenshot shows a web browser window with multiple tabs open at the top. The main content area displays the 'Downloads' section for Altair AI products. A yellow highlight bar is drawn across the page, starting from the 'Downloads' tab in the navigation bar and extending down to the 'AI Hub 2024.0' section.

**Downloads**  
Click on an Altair AI product of your choice to download it.

**AI Studio 2024.0**  
Click on your operating system to start the download:

- Windows** 64bit
- Mac**
- Linux** 64bit  
Requires: Java 11

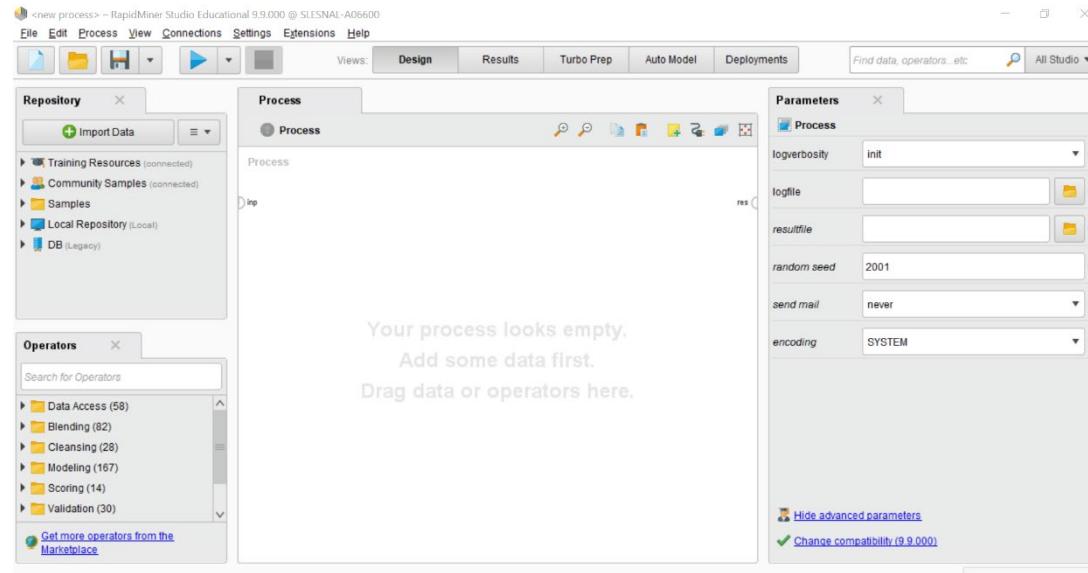
**AI Hub 2024.0**  
You need to log into your account to download AI Hub.  
[Sign in](#) or [Register if you don't have an account](#)

**Radoop 2024.0**  
You need to log into your account to download Radoop.  
[Sign in](#) or [Register if you don't have an account](#)

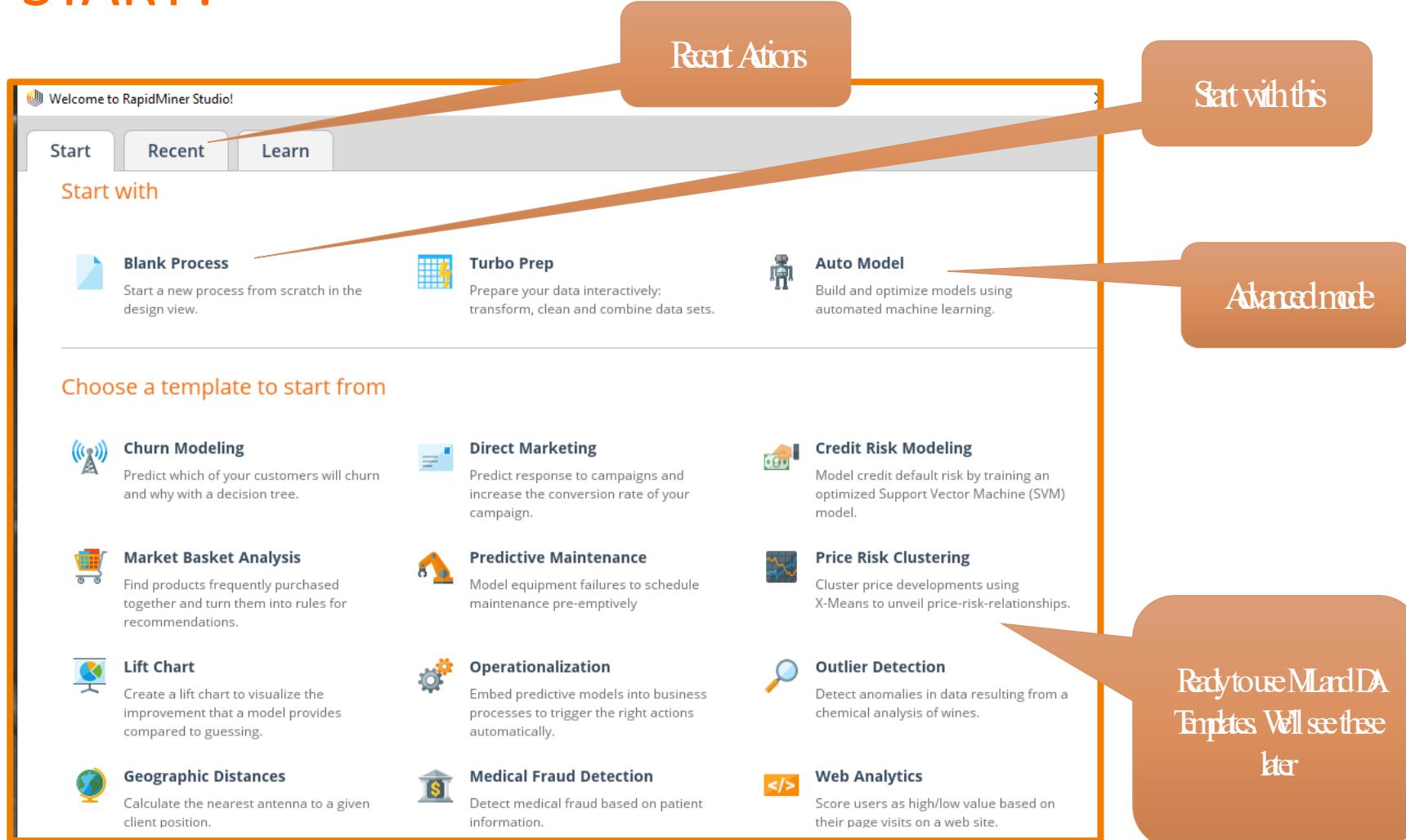
At the bottom of the screen, the Windows taskbar is visible, showing the Start button, a search bar with the text 'Type here to search', and various pinned icons for applications like File Explorer, Mail, and Edge. The system tray shows the date (16-Jun-24), time (17:16), battery level (30°C), and network status.

# HOW TO START?

- Mind your OS
- Install it
  - The basic **RapidMiner Studio** is free.
- Run it !!!
- [www.rapidminer.com](http://www.rapidminer.com)



# HOW TO START?





Views:

Design

Results

Turbo Prep

Auto Model

Deployments

Find data, operators...etc



A

**Repository**[Import Data](#)

- ▶ **Training Resources** (connected)
- ▶ **Community Samples** (connected)
- ▶ **Samples**
- ▶ **Local Repository** (Locally available)
- ▶ **DB Jdbc** (available)

# HOW TO START?

**Operators**

Search for Operators

- ▶ **Data Access** (58)
- ▶ **Blending** (82)
- ▶ **Cleansing** (28)
- ▶ **Modeling** (167)
- ▶ **Scoring** (14)
- ▶ **Validation** (30)

[Get more operators from the Marketplace](#)**Process****Process**

Your process looks empty.

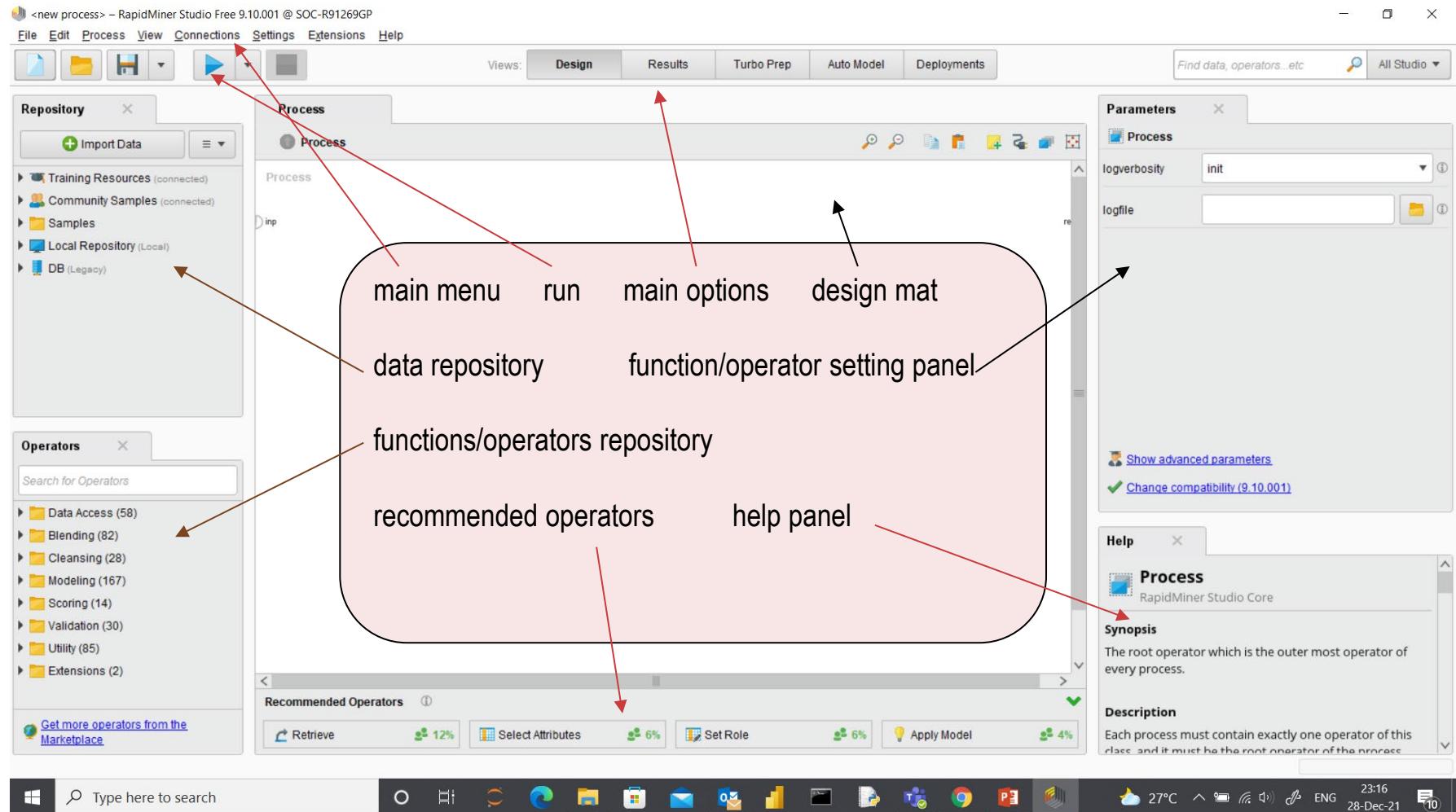
Add some data first.

Drag data or operators here.

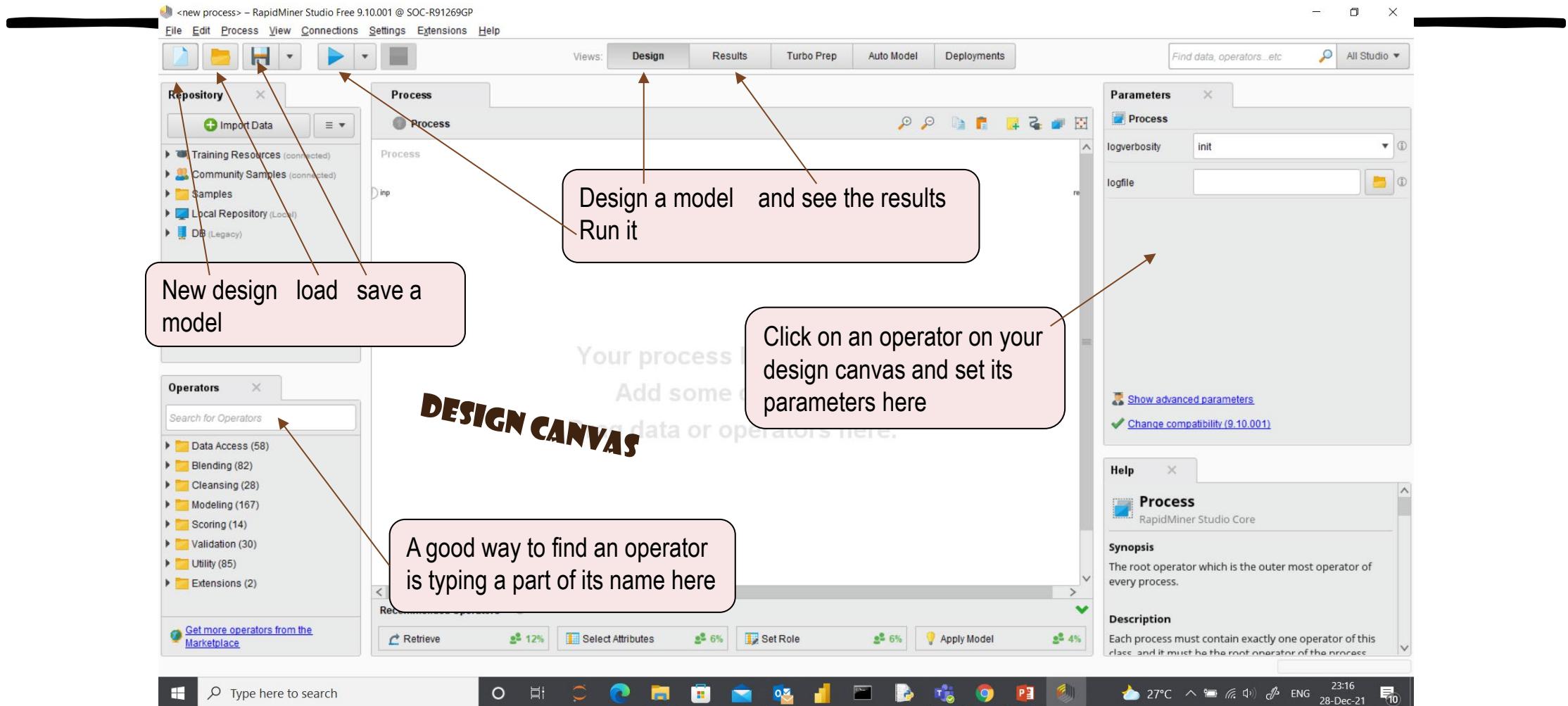
RAPIDMINER PAGE, WHAT DO SET IT

**Parameters****Process**logverbosity logfile resultfile random seed send mail encoding [Hide advanced parameters](#)[Change compatibility \(9.9.000\)](#)

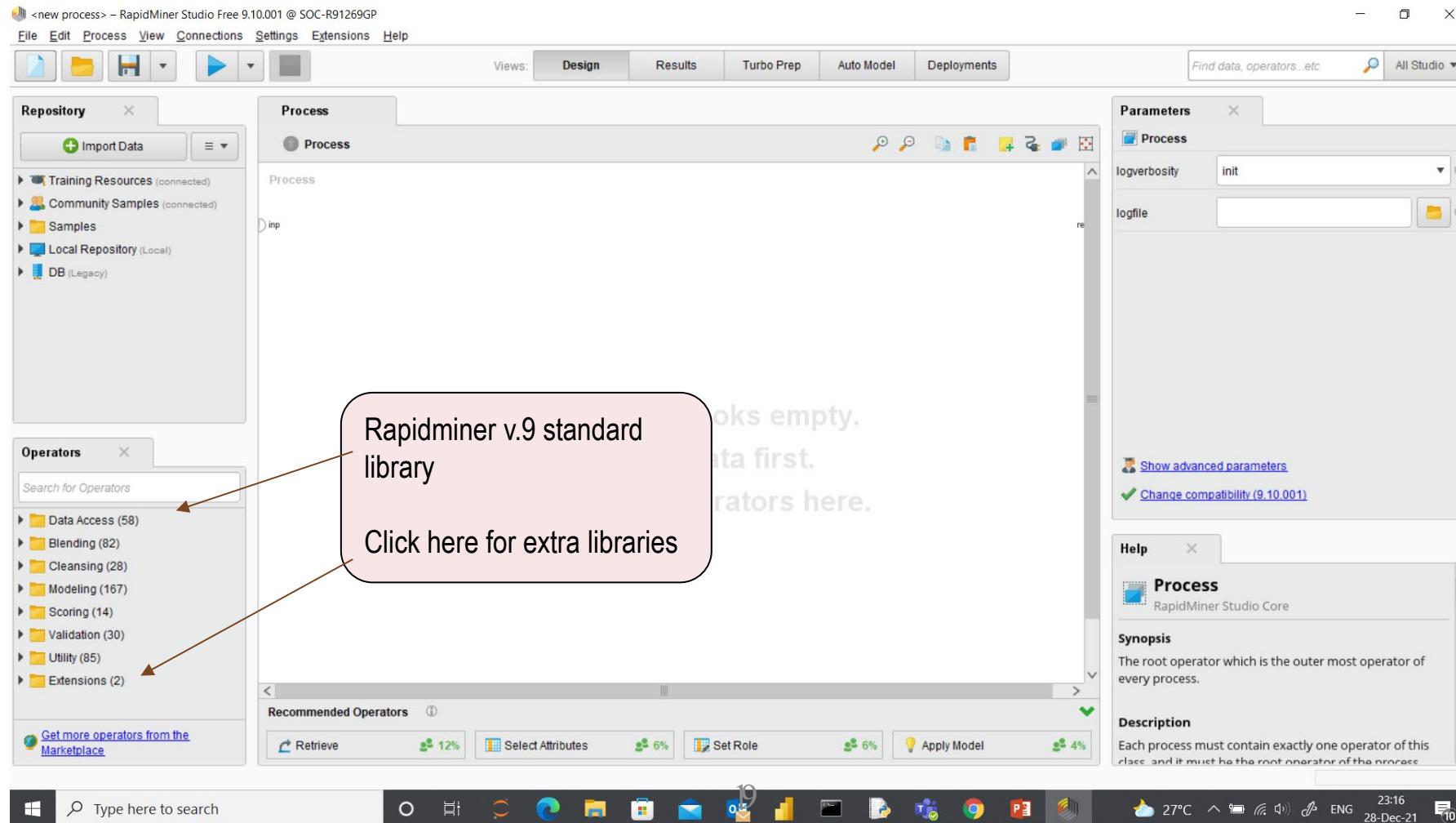
# RAPIDMINER MAIN PAGE



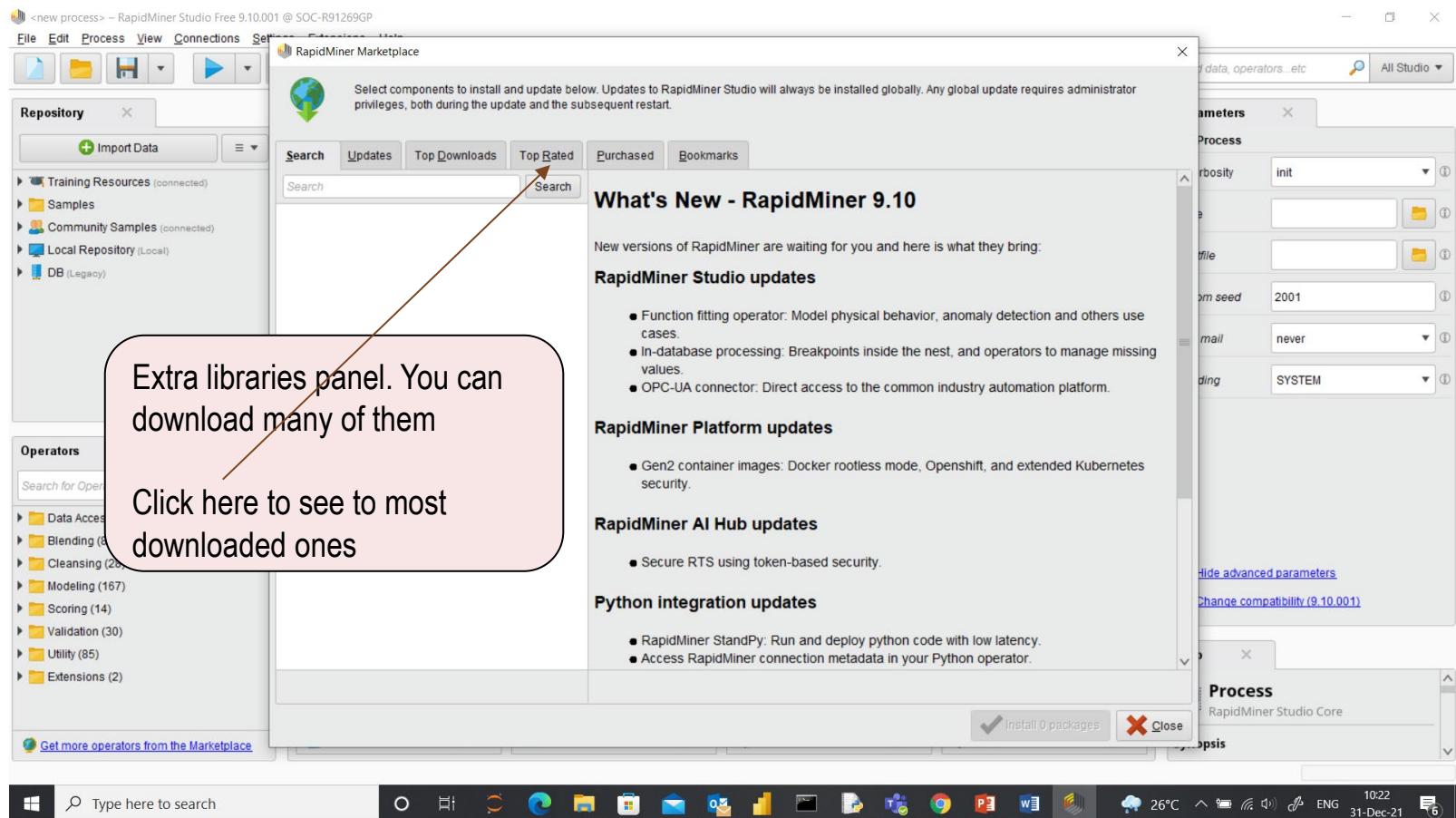
# RapidMiner Main page



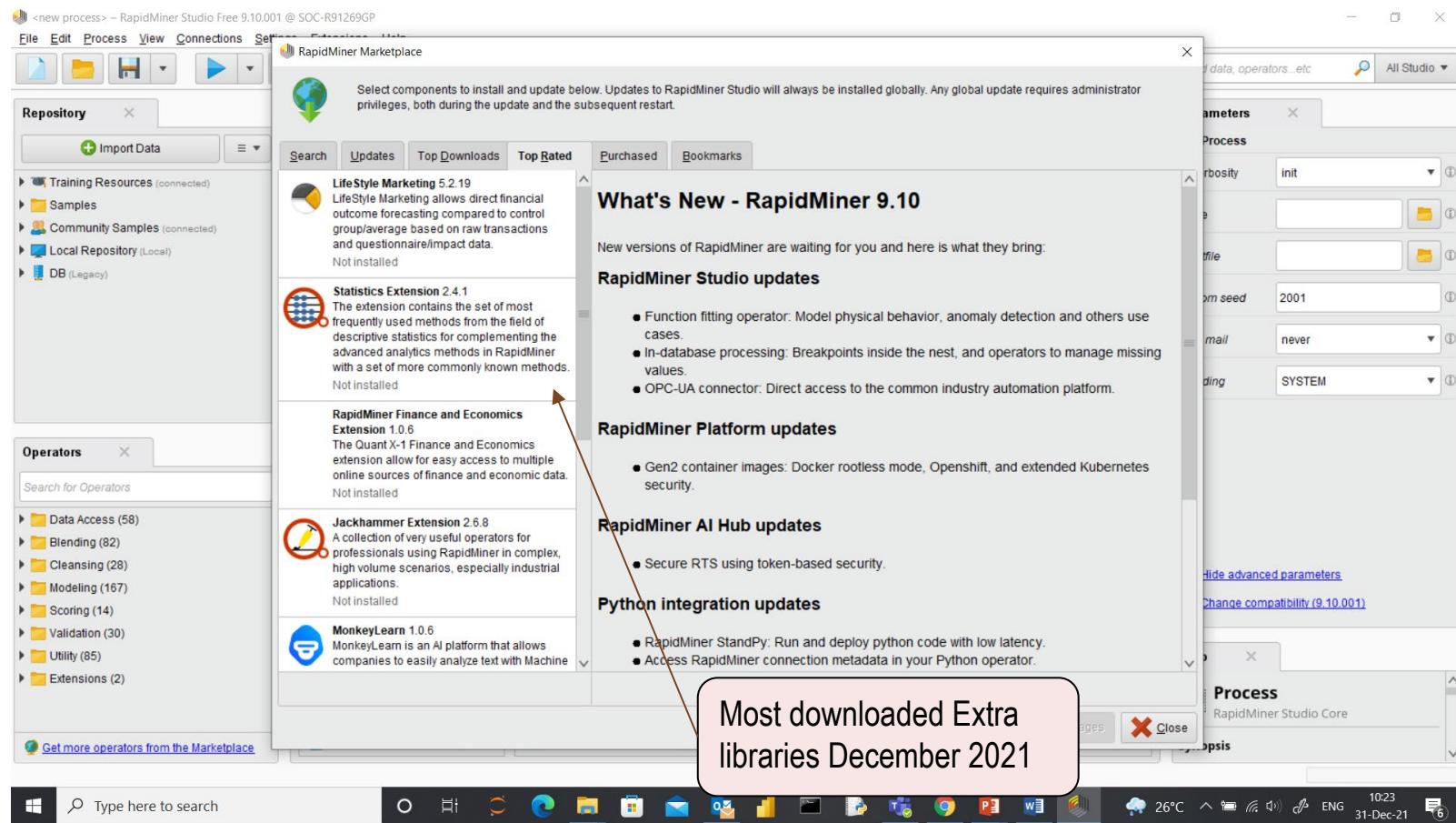
# Libraries



# Libraries



# Libraries





# Data Preprocessing

---

# Data Preprocessing Example

---

- Datafile celebs1-bad-25.csv
- Python notebook preprocessing-25-lipynb
- Goals
  - Handling missing values by removing
  - Handling missing values by replacement
  - Datamalization
  - Dtydata handling

# Data file

name	height	weight	gender	DB
Adriana Lima	178	115f		Jun-81
Alicia Keys	167.64	120f		Jan-81
Angelina Jolie	172.7	122f		Jun-75
Ana Konkova	172.7	120f		Jun-81
Ashley Judd	170.2	125f		Apr-68
Amber Heard	155	113f		Sep-84
Beyoncé	174	128f		Jun-85
Brandy Norwood	167.6	129f		Sep-81
Celine Dion	172.7	f		Aug-87
Constance Shulman	160	113f		Apr-72



# The End...

Ttoid 1