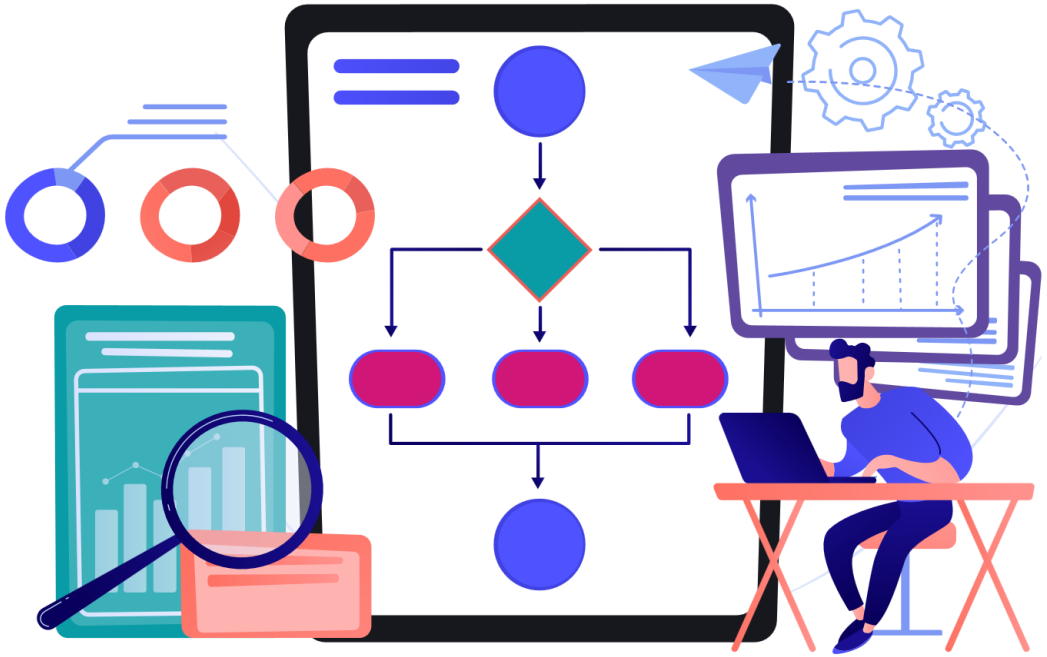
blueprism®

blueprism
Process Intelligence
powered by ABBYY Timeline

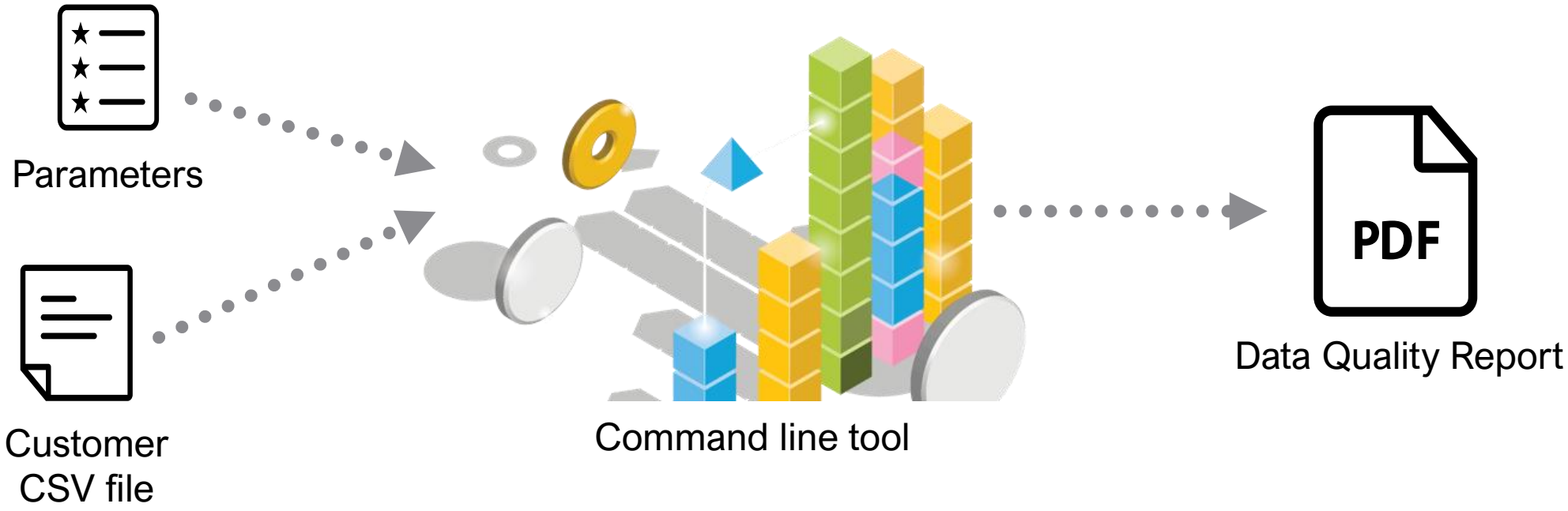


BPPI

Data Quality Kit for BPPI (DRAFT)

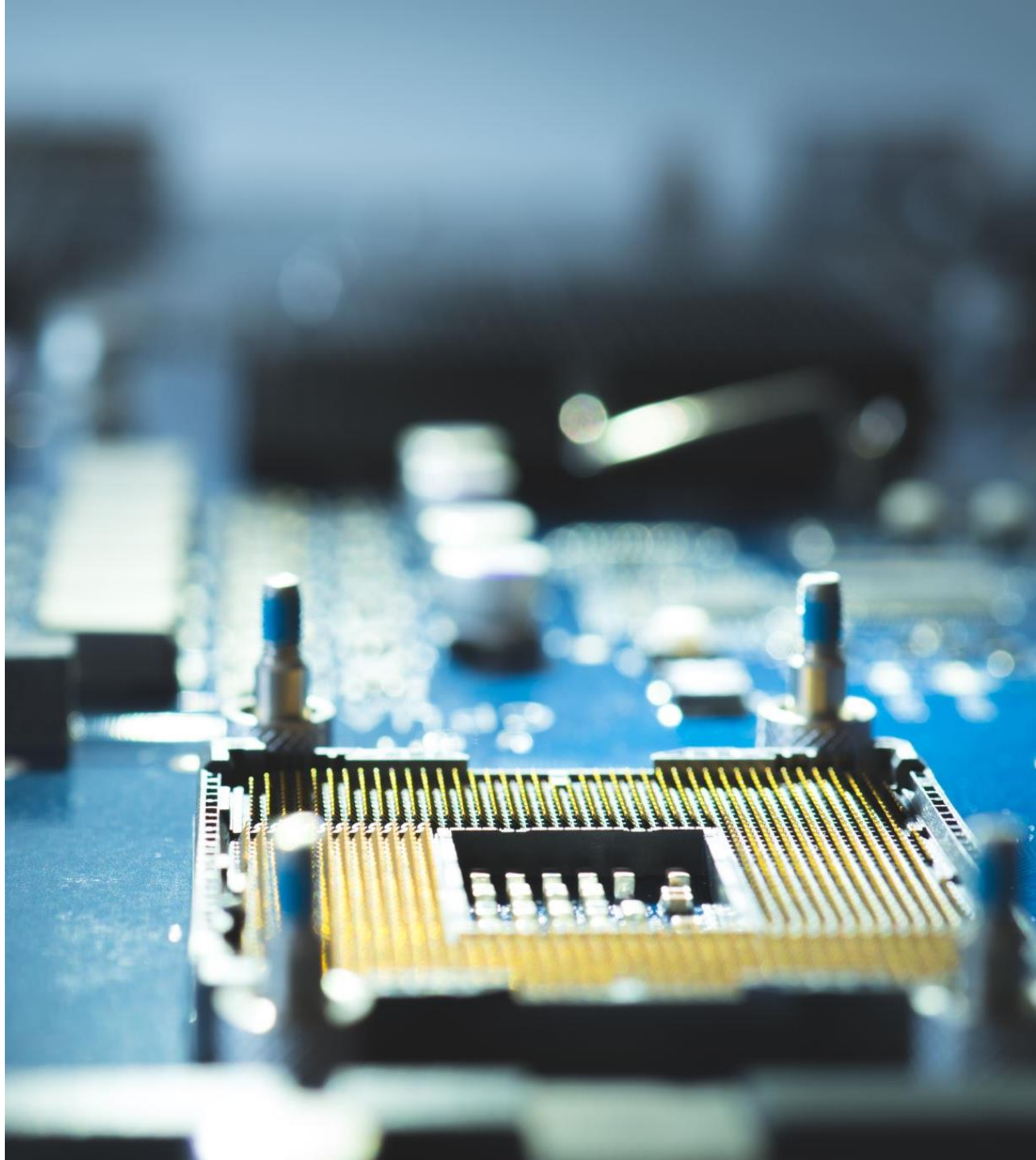
What is it exactly ?

- This DQ Kits helps to evaluate the datasets given by a customer
- Can only evaluate 1 dataset at a time
- Provides a pdf report as result
  - This report can be used during the BPPI data workshops



Pre-Requisites

- One Dataset provided in CSV format
- KNIME v4.6.1 installed (free download GPL-3 license)
  - https://www.knime.com/downloads
- BPPI Data Quality Kit (zip) installation:
  - Just by importing the knar file into KNIME
  - Adjust the command line parameters



How to use the kit ?

By simply running a command line on Windows (cmd.exe)

Example:

"C:\Program Files\KNIME\knime.exe" --launcher.suppressErrors -reset -nosave -consolelog -nosplash -application org.knime.product.KNIME_BATCH_APPLICATION -workflowDir="C:\knime-wk\BPPI Toolbox\BPPI_DataAnalysis_BuildReport" -workflow.variable="BPPI_OutputPath","C:\\knime-wk\\BPPI Toolbox","String" -workflow.variable="file","C:\\knime-wk\\BPPI Toolbox\\data.csv","String" -workflow.variable="TIMELINEID_Column","TimelineID","String" -workflow.variable="TIMESTAMP_Column","Date","String" -workflow.variable="EVENTID_Column","Event","String“ -workflow.variable=“delimiter“,“;“,"String“

How to use the kit ?

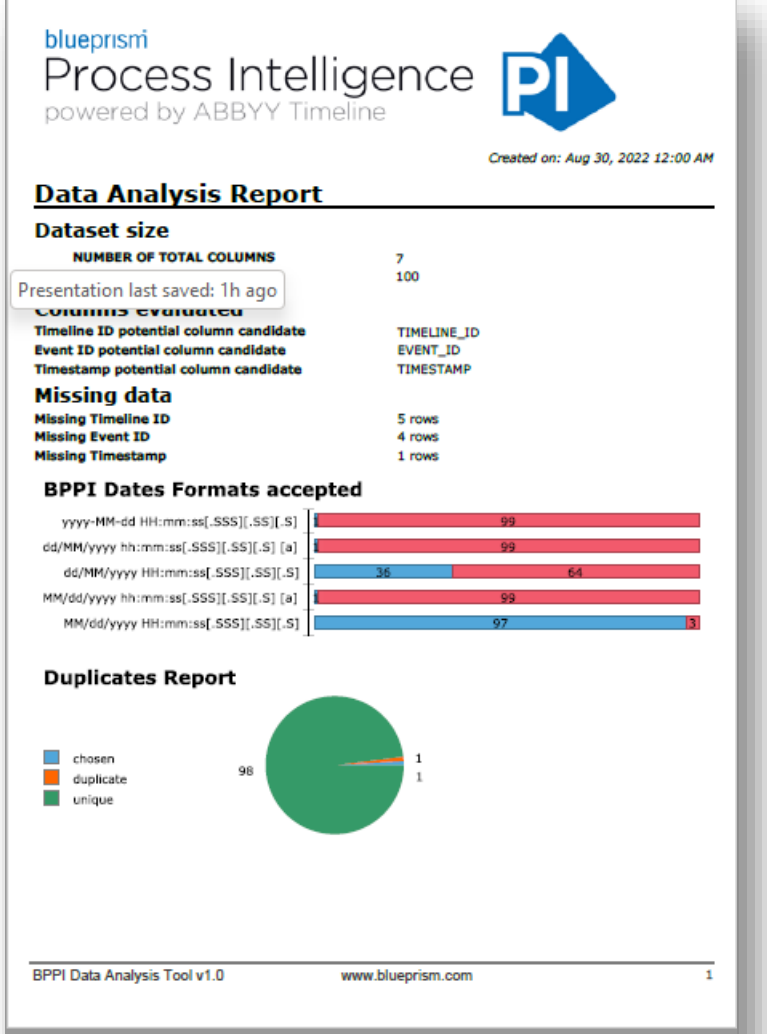
Command line parameters:

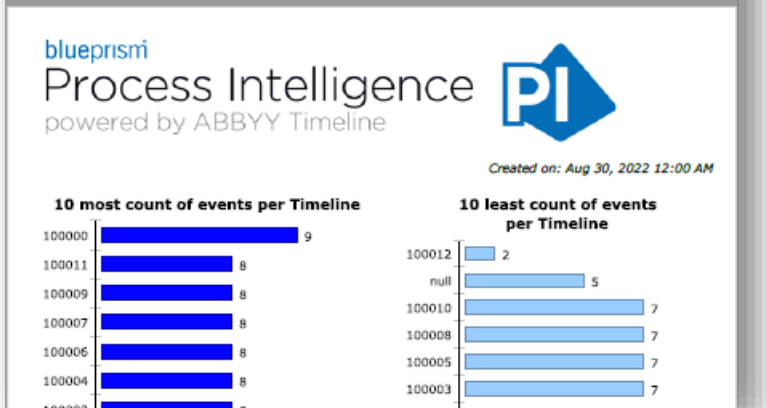
You can change these parameters in the command line:

- *BPPI_OutputPath : Put here the path where you want to get the reports (results)*
- *file : datasource file (CSV format)*
- *delimiter : CSV delimiter (comma by default)*
- *TIMELINEID_Column: Name of the TIMELINE_ID column candidate*
- *EVENTID_Column: Name of the EVENT_ID column candidate*
- *TIMESTAMP_Column: Name of the TIMESTAMP column candidate*

Report description

- Dataset Header & Size
- Columns candidates (evaluated)
- Missing Data
- Date Format checks
- Duplicates checks
- Most & Least count of events / Timeline
- Statistics on Count(Event) per Timeline
- Events Freq. distribution
- Number of rows / Timeline





Dataset Header

blueprism
Process Intelligence
powered by ABBYY Timeline

Created on: Aug 30, 2022 12:00 AM

Data Analysis Report

Dataset size

NUMBER OF TOTAL COLUMNS	7
NUMBER OF TOTAL ROWS	100

Columns evaluated

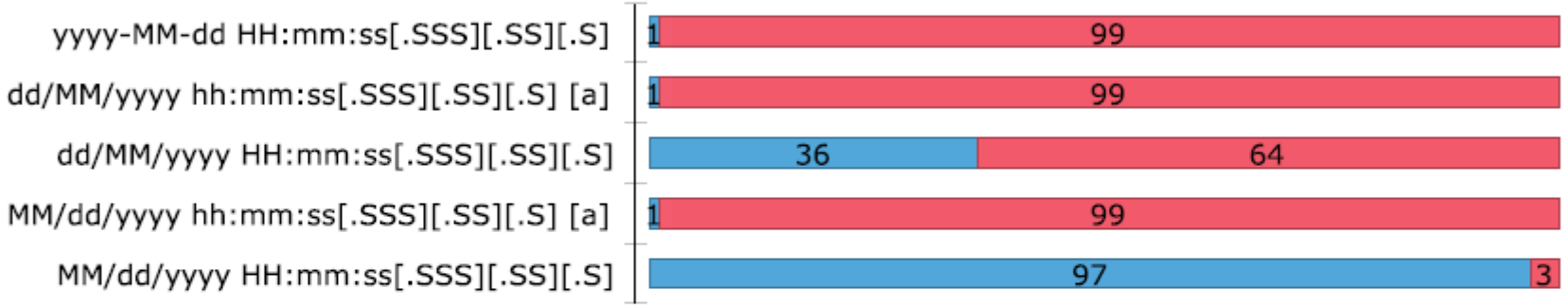
Timeline ID potential column candidate	TIMELINE_ID
Event ID potential column candidate	EVENT_ID
Timestamp potential column candidate	TIMESTAMP

Data Quality informations

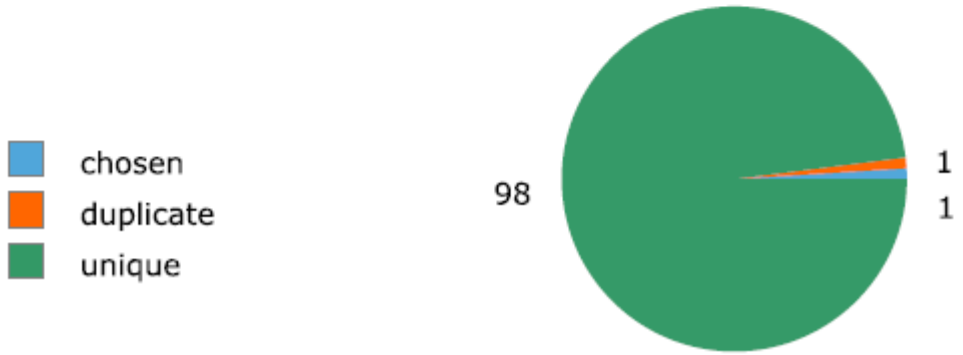
Missing data

Missing Timeline ID	5 rows
Missing Event ID	4 rows
Missing Timestamp	1 rows

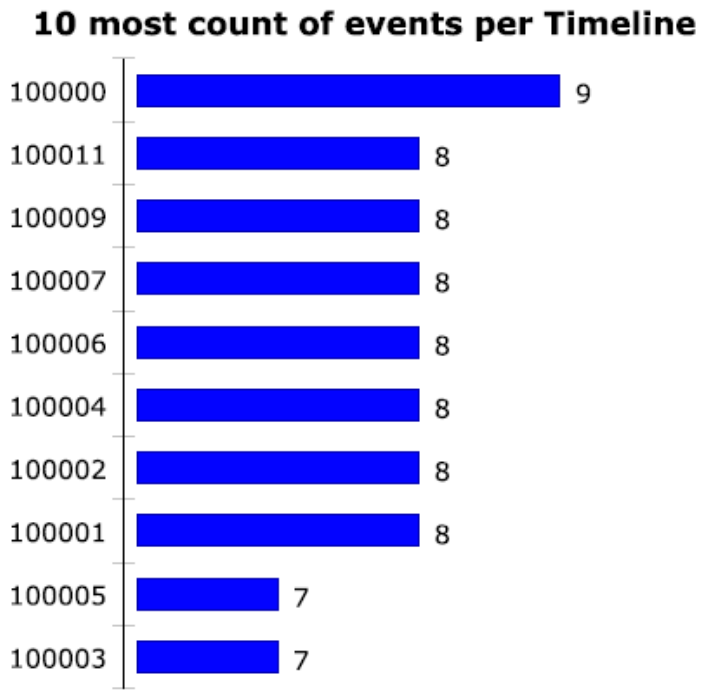
BPPI Dates Formats accepted

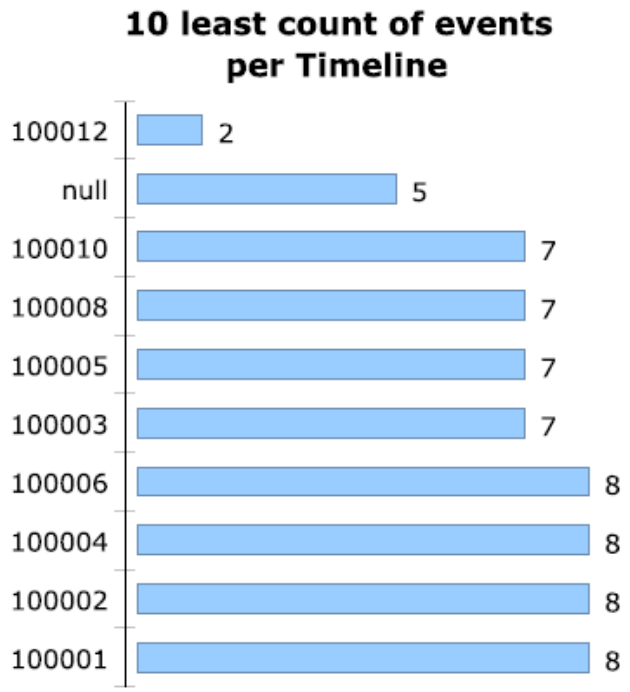


Duplicates Report



EVENT_ID details



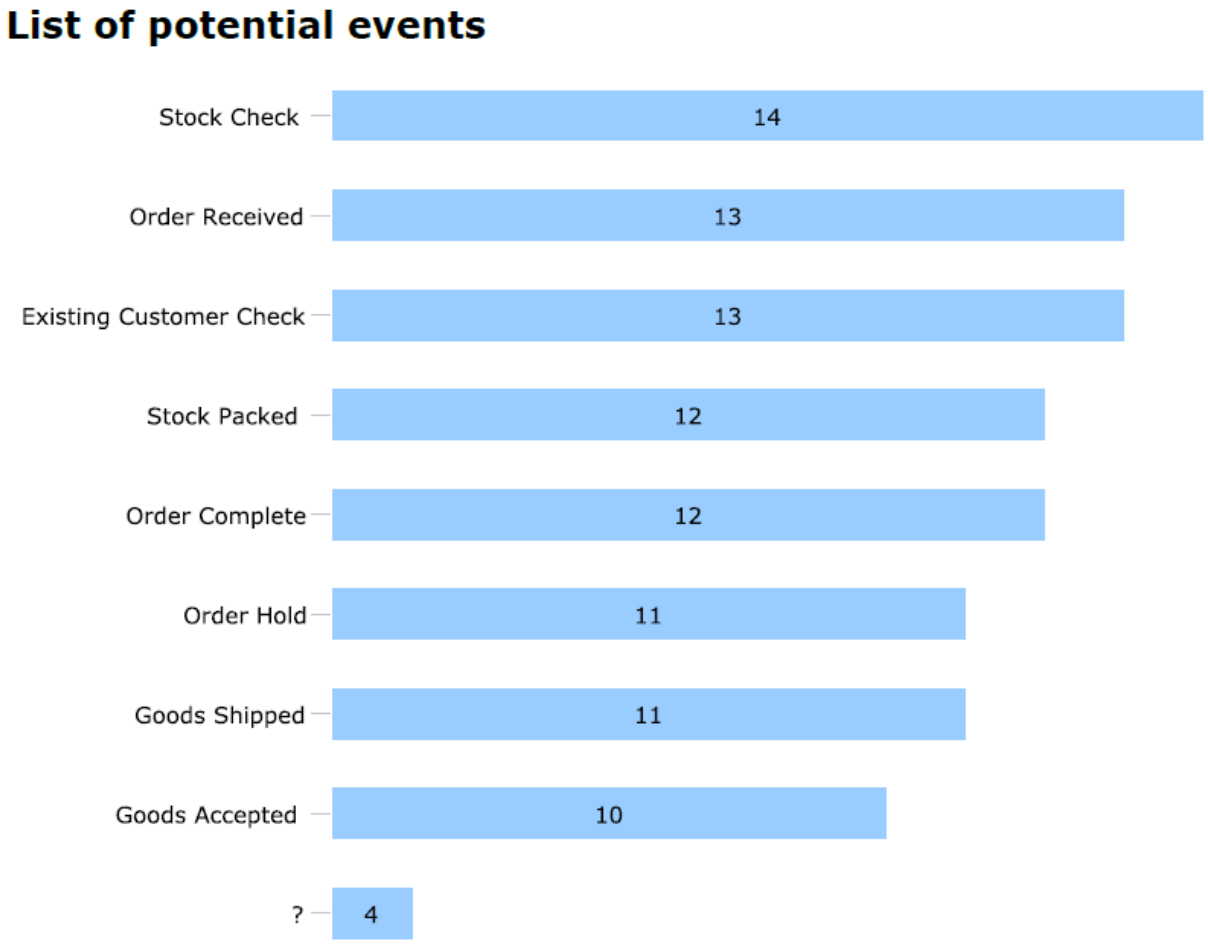


These two reports show the maximum & minimum number of events counted per timeline.

Event ID Count / Timeline ID

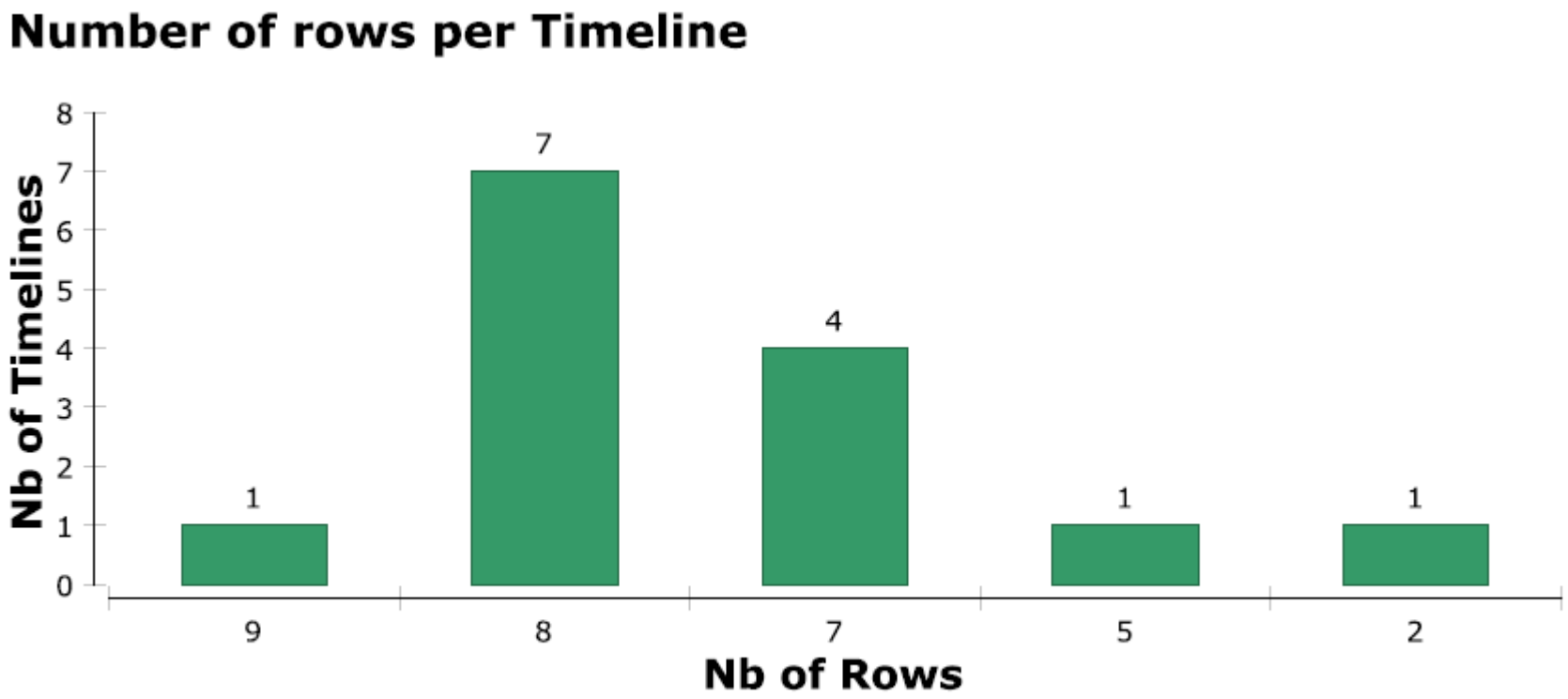
Min	2
Max	9
Mean	7.142857142857142
Std. deviation	1.747840111378914
Variance	3.0549450549450543
Median	8

Events Freq. Distribution



This report displays all the differents events and their frequency distribution.

Event – Timeline consistency



This report counts the number of rows (can be different as the number of events) per timelines and group them.

blueprism®

A Digital Workforce for Every Enterprise