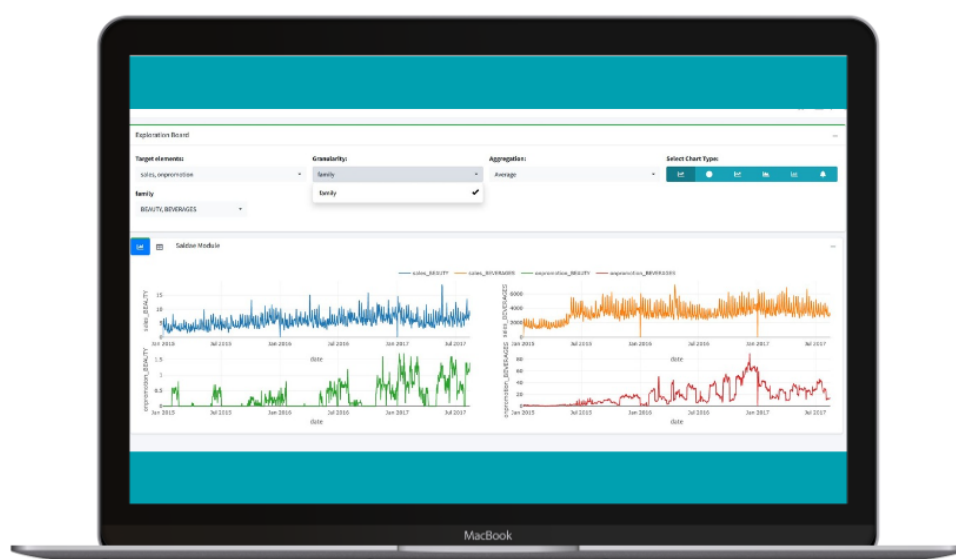




FairViewer by FairAnalytics: Documentation



Email : contact@fair-analytics.com

FairViewer by **FairAnalytics** is a code-free & easy-to-use web-based platform, which provides a powerful tool for your [time series](#) data **analysis** and **forecasting**. It is designed following high security recommendations and it's fully GPRD compliant.

With an attractive interface, secure access and very easy data loading capabilities using multiple connectors. Our solution goes beyond classical solutions which are already on the market, which require prior knowledge in data analysis and AI to perform Data exploration, Visualization, AI based Analytics and reporting. Below is a set of feature integrated in **FairViewer** :

1. Data handling:

a. Data Connectors :

Our software's strength is its ability to import data from a variety of data sources using our built-in connectors (.csv, .txt, .xlsx, SQL, NoSQL Databases, Cloud storage S3, ...) and to perform very fluid data wrangling/manipulation.

Your dataset should include a variable(column) containing **date/time** information in one of the following formats (YYYY/MM/DD,YYYY-MM-DD,YYYYMM/DD,...).

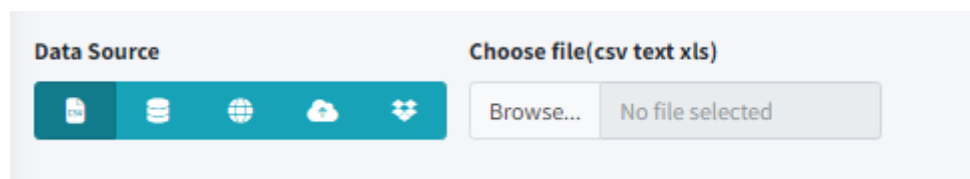


Fig 1 : Dataset Loader

b. Data Types:

An automatic detection of data types is performed to split the dataset into numerical & categorical.

here are some useful details:

- **Numerical Data** : are used for plots, analytics & statistics.
- **Categorical Data** : used for more detailed and specific data manipulation.

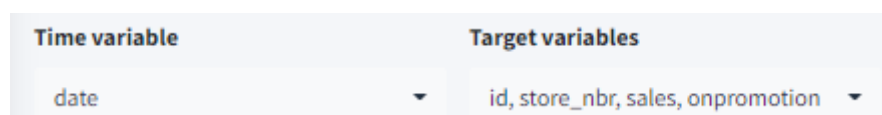


Fig 2 : Dataset variables

c. Data Quality :

In the backend, automatic missing values' detection is integrated. It will return the percentage of complete data. A time-based interpolation is then performed for numerical values to complete the dataset & improve analytics quality.

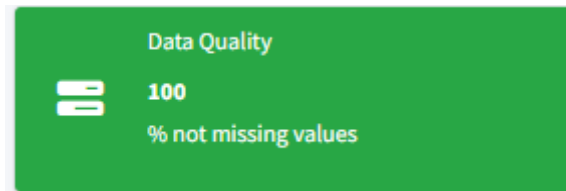


Fig 3 : Dataset Quality pourcentage

d. Data Frequency :

Data frequency detection algorithm is implemented to recognize the time frequency of your dataset(minutes, 1/4 hours, 1/2 hours, hours, days, weeks, quarters, years)

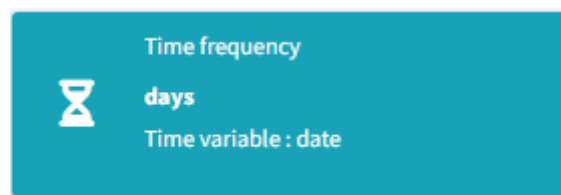


Fig 4 : Dataset Frequency

e. Statistics :

The most relevant statistical parameters are interactively calculated :

- **Mean** : is the sum of all of the numbers divided by the number of numbers.
- **Standard deviation** : is the degree of dispersion or the scatter of the data points relative to its mean.
- **Dataset distribution** : the spread of the data which shows all possible values or intervals of the data and how they occur.
- **Skewness** : is a measure of the asymmetry of the dataset distribution.
- **Kurtoisis** : is a measure of the tailedness of the dataset distribution.

The table below shows an example of all the estimated parameters for every numerical variable in the dataset :

Start Date	End Date	variables	outliers_cnt	n	na	mean	sd	se_mean	IQR	skewness	kurtosis
01/01/2015	08/15/2017	id	0	315480	0	2149263.10	491785.63	875.57	851795.50	0.00	-1.20
01/01/2015	08/15/2017	store_nbr	0	315480	0	5.50	2.87	0.01	5.00	0.00	-1.22
01/01/2015	08/15/2017	sales	43975	315480	0	496.76	1308.46	2.33	321.48	7.55	303.22
01/01/2015	08/15/2017	onpromotion	64289	315480	0	4.54	16.10	0.03	1.00	7.51	76.27

Fig 5 : Dataset Statistics

2. Data Visualisation :

We offer a complete tool to perform interactive visualizations and be able to represent the information graphically. We take into consideration different categorical data to create granularities & to give you the possibility to visualize every simple case.

- Exploration:** gives a matrix of graphs of all the variables at the same time to have a general view of the dataset. Here are the options that you can modify to draw different visualization plots :
 - Select Parameters/Granularity (matrix of plots is incremented automatically).
 - Change aggregation parameters from : sum, max, min, mean or median.
 - Change plotting type : line graph, scatter, linked scatter, data distribution.

Here is the illustration of the result :

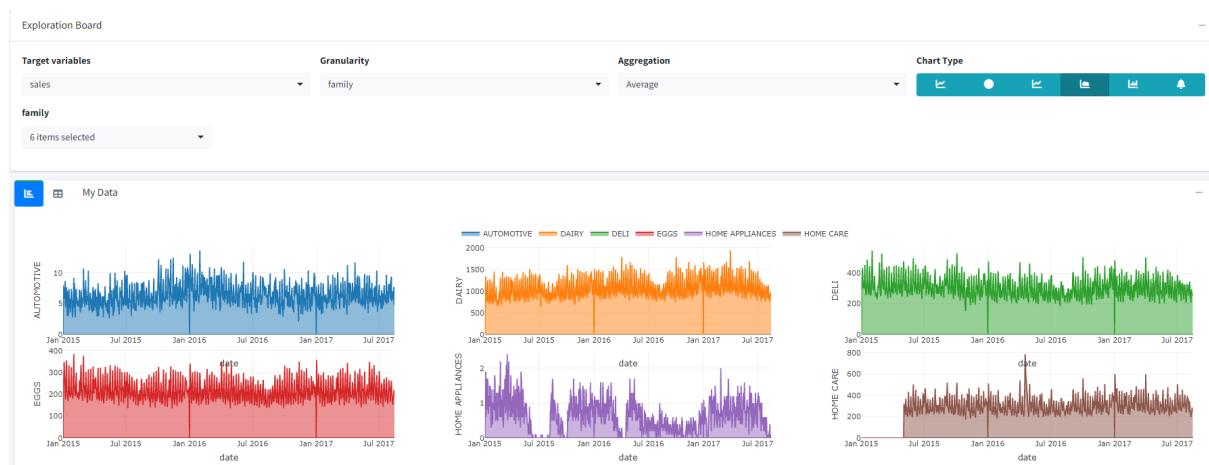


Fig 6 : General Visualisation of dataset

- b. **Aggregation:** This functionality has exactly the same characteristics as visualization but you have the ability to add an aggregation time over days, weeks, months or years. here is the list of the options:
- Select Parameters/Granularity (matrix of plots is incremented automatically).
 - Change aggregation parameters from : sum, max, min, mean or median.
 - Change plotting type : line graph, scatter, linked scatter, data distribution.
 - Select aggregation parameter from : week, month (time aggregation depends on your data time frequency).

The illustration below shows the result :

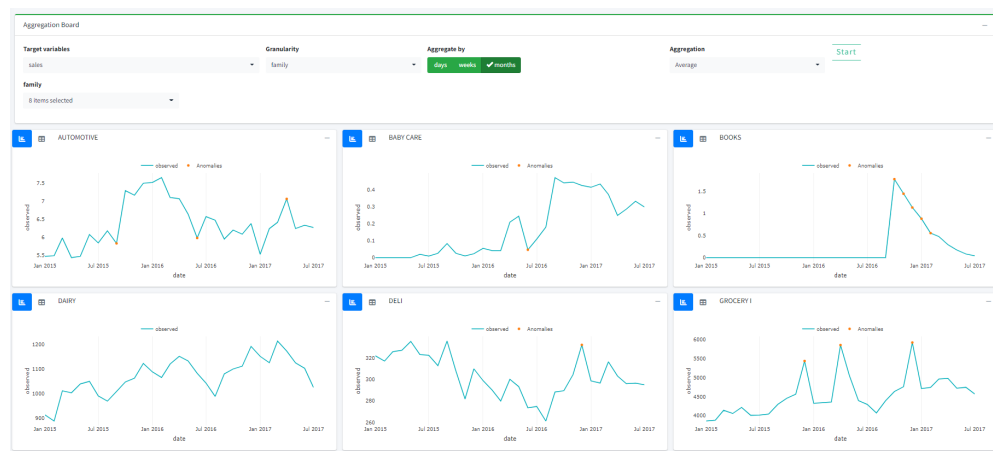


Fig 7: Monthly aggregated dataset visualization

- c. **Growth rate:** This functionality will allow you to visualize the growth rate of your dataset by selecting a suitable rate (days, weeks or months). The illustration below shows the result :

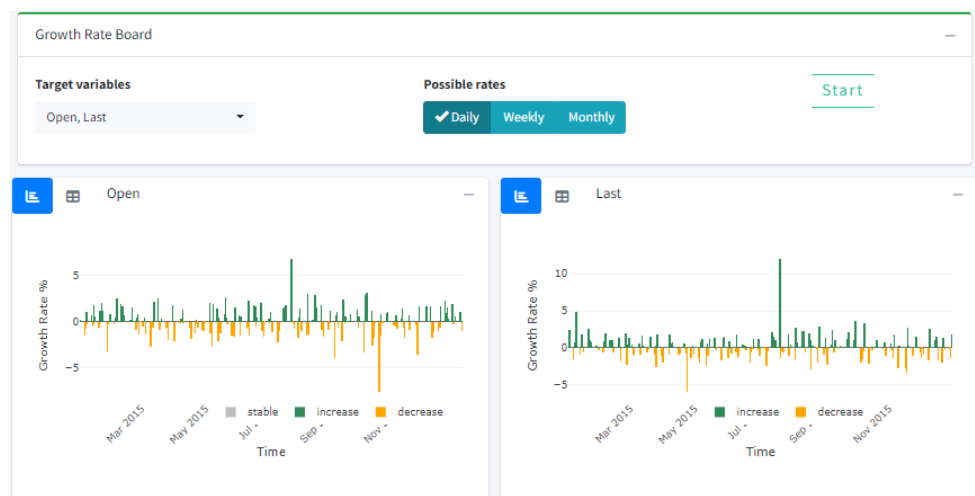


Fig 8: Growth rate example illustration

3. Analytics :

Our solution is smart and very efficient for data analysis. We have designed, implemented & integrated the most useful AI & data-based algorithms to perform a complete time series analysis.

It aims to extract most useful features and parameters from your datasets, shows the hidden interdependencies and connections such as correlations & causality between your features and KPIs.

We can help you to recycle your data and to explore for better decision making. We gathered all our knowledge to create an algorithm which will show you how to explore your data ? Which part of the dataset should you consider ? What are the most relevant features ? and even to highlight the correlations & causality between the selected parameters?

3.1. Features importance/selection :

a. definition : it's the process of selecting a subset from the dataset to make the models more relevant. It can be used to extract hidden interconnection between your dataset variables.

b. Tutorial : Based on multiple data analysis algorithms, we designed a solution for a features importance dedicated for every target variable; i.e, we show the most relevant features for your target variable. The user has to be selected and the algorithm will look for potential correlated & connected variables.

Here are the instructions :

- In the **FairViewer** tool, go to “**AI & Analytics**” tab.
- Select “**Interdependency**”.

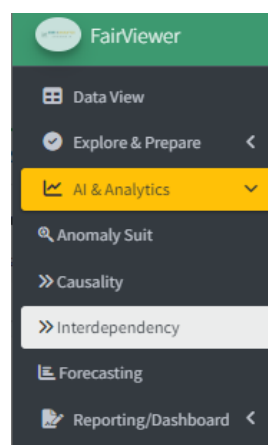


Fig 9 : Interdependency functionality

- Select the target variable for the analysis.
- Click on “**Start**”.

The figure below, shows the interface with the instructions :

The screenshot shows a web interface titled 'Interdependency Board'. Below the title, there is a section labeled 'Target Variable' with a dropdown menu currently showing 'Prev_Close'. To the right of this dropdown is a green button labeled 'Start'.

Fig 10 : Interdependency Tool

This functionality will frisk your dataset in all the ways to extract potential correlated features & interdependencies between parameters. The result is displayed using bins and it's illustrated in the figure below :

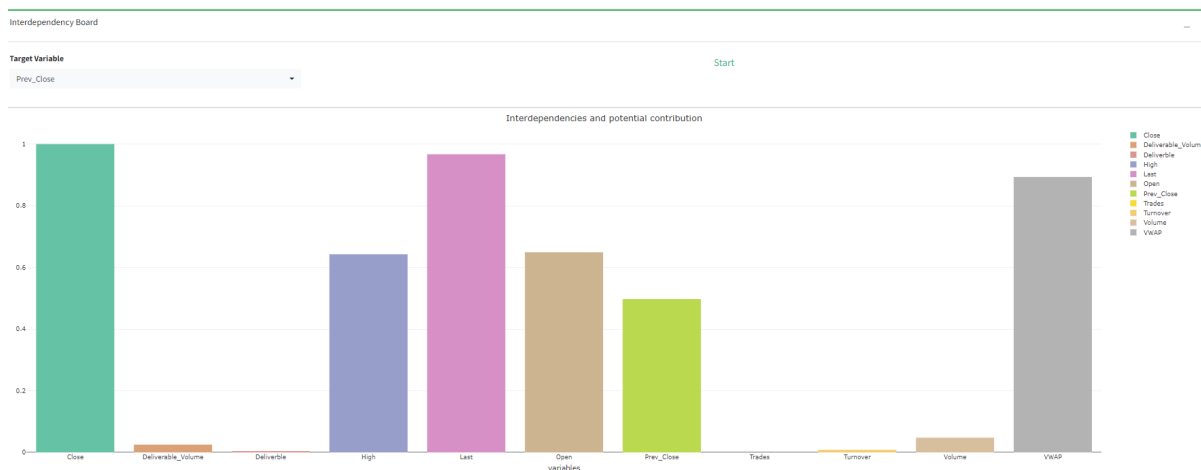


Fig 11: Interdependence KPIs visualization

3.2. Anomaly detections :

a. Definition : Anomalies detection is the identification of rare items, events or observations which deviate significantly from the majority of the data and do not conform to a well defined notion of normal behavior.

b. Tutorial: based on one of the most used unsupervised AI algorithms, we integrate outlier detection for an automatic detection of undesirable events & values/points in your dataset/variable. This function will allow you to know the dates & values of these events, remove outliers, make your dataset clean and your model more robust & efficient.

Here are the instructions :

- In the **FairViewer** tool, go to the “**AI & Analytics**” tab.
- Select “**Anomaly suit**”.

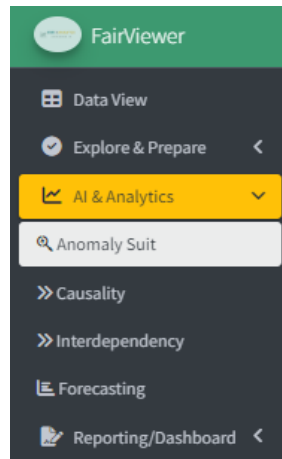


Fig 12: Anomaly detection Functionality

- Select the target variable for anomalies detection.
- Click on “**Start**”.

The figure below, shows the interface with the instructions :

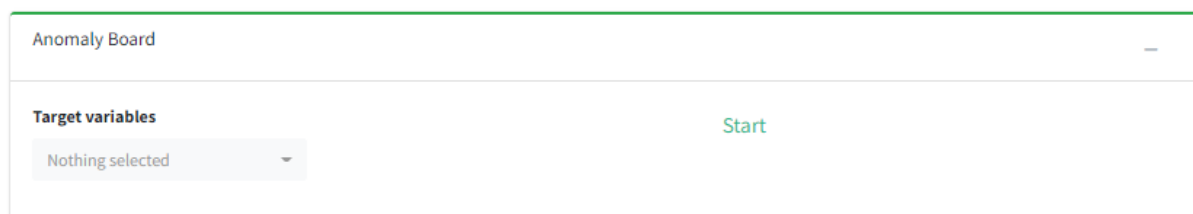


Fig 13: Anomaly detection Tool

This module returns the results in 2 forms :

- **Table** : The full variable table where all the detected outliers with (dates & frequency) are highlighted in orange/maroon color(upper/lower outlier).
Fig : Table with normal data points & outliers
- **Plot**: the orange points show all the detected outlier points where the behavior of the dataset is supposed to be undesirable.

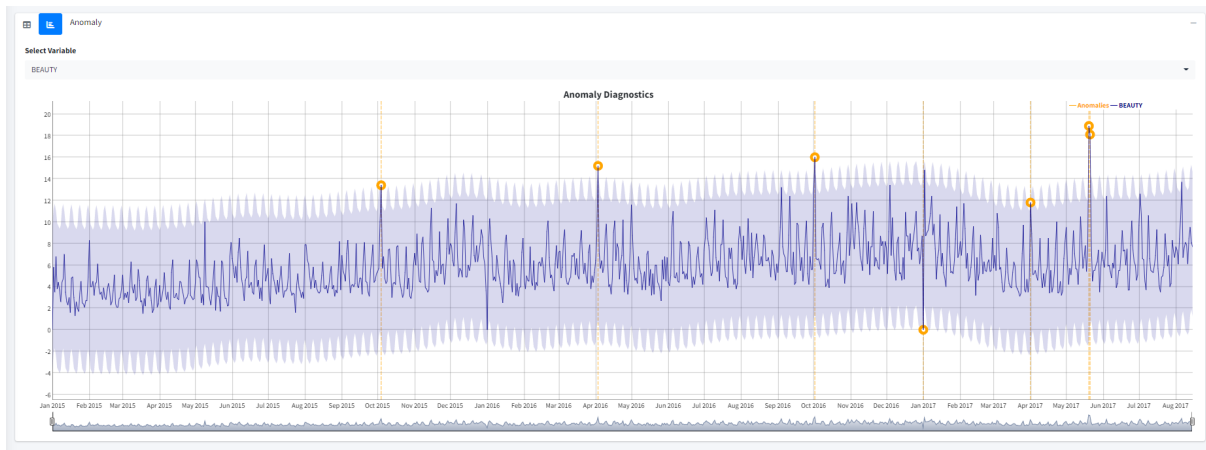


Fig 13 : Anomalies detection illustration with dates

3.3. Causality :

a. Definition : Causality is a relationship between two events, or variables, in which one event or process causes an effect on the other event or process.

b. Tutorial : This functionality requires 2 steps to work. Firstly , it needs to perform dataset dimensionality reduction & clustering; i.e, the user has to select a number of clusters. Then, the method estimates the interconnections and similarities between variables.

Here are the instructions :

- In the **“FairView”** app, go to **“AI & Analytics”** tab
- select **“Causality”**.

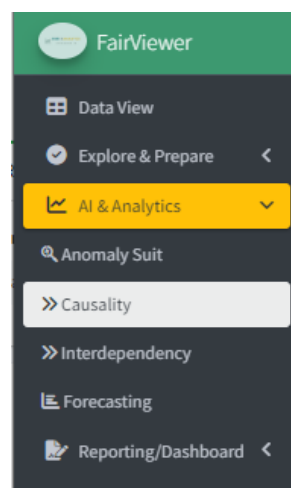


Fig 14: Causality Functionality

- Select the target variables (for full dataset analysis, select all variables).

- Enter Number of clusters
- Click on “**Start**”.

The results of the module is given in 2 forms :

- **General Plot** : the output is a reactive 2D representation of the variable inter-dependencies (correlation, causation).

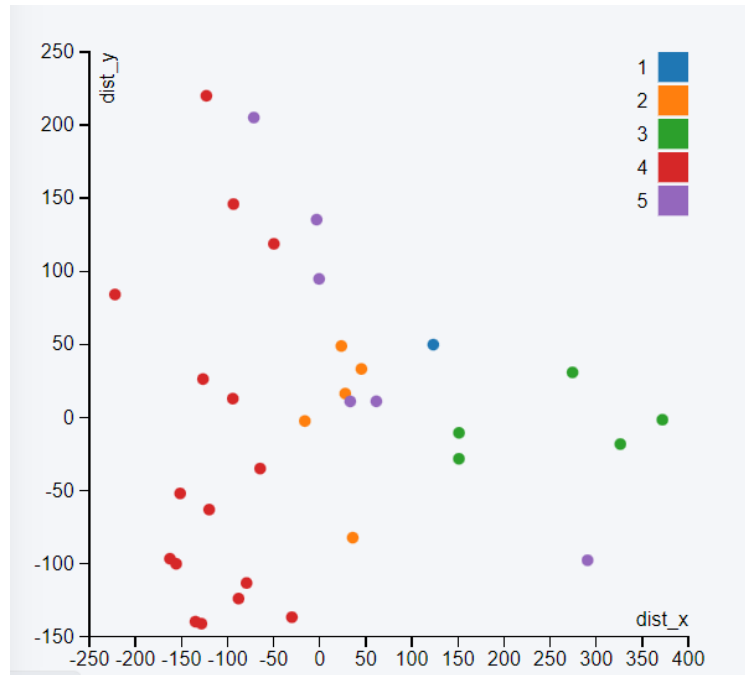


Fig 15 : Example of Clustering

- **Causality inter-features** : when we select 2 or more points close from each-other and with the same color, the algorithm has detected significant inter-dependency/connection :

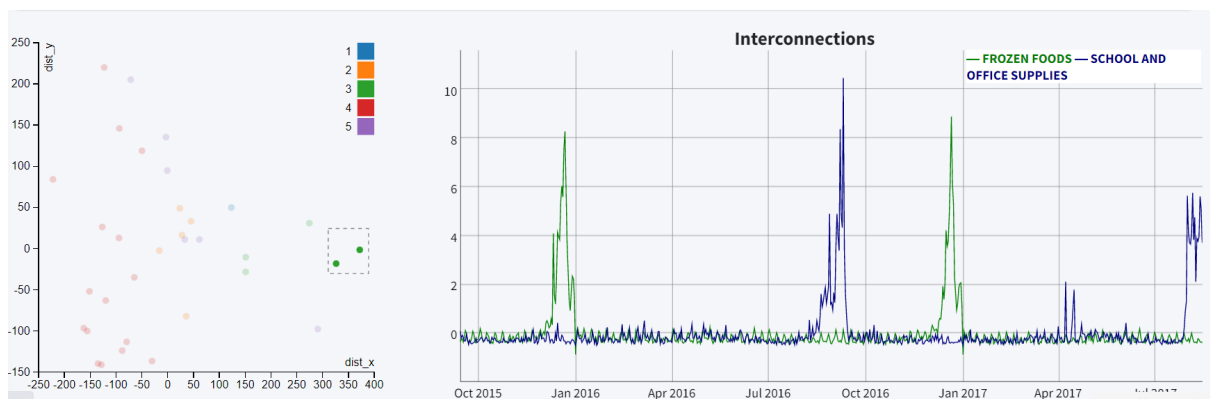


Fig 16 : Example of Causality

3.4. Forecasting :

a. Definition : is the process of making predictions based on historical and present data. Many statistical methods (ARMA,ARIMA,MA...) and machine learning algorithms (XGboost, LSTM ...) are well known to provide very good results for time series forecasting.

b. Tutorial : This functionality is implemented to give a very strong tool to make forecasts on future demand/sales/arrivals/.. for better decision making. In addition it provides a 95% confidence interval.

For quality concerns, the forecast horizon is fixed to a maximum of 25% of the length of historical data (for 1 year dataset => it will return max 3 month forecast results).

Here are the instructions :

- In the “**FairViewer**” app, go to “**AI & Analytics**” tab
- select “**Forecasting**”.

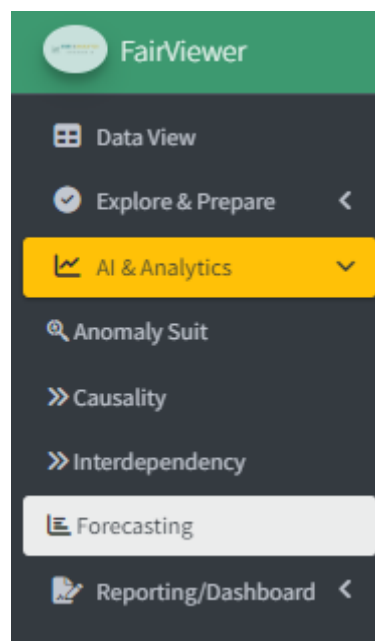


Fig 17 : Forecasting Functionality tool

- Select the target variable for forecasting.
- Activate/deactivate anomalies detections.
- Click on “**Start Predictions**”.

We can see in the figure below, forecasting results and the corresponding confidence intervals :

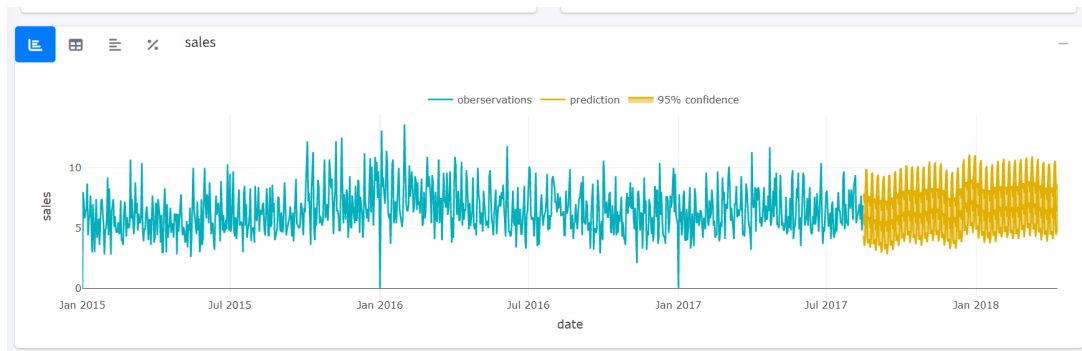


Fig 18 : Forecasting graph

4. Dashboard & Reporting :

As a next step once you have performed your exploratory data analysis, you export/publish the outcome as a detailed interactive report. In the “**FairViewer**” app, go to the “**Reporting/Dashboard**” tab, click on **Generate and publish**.

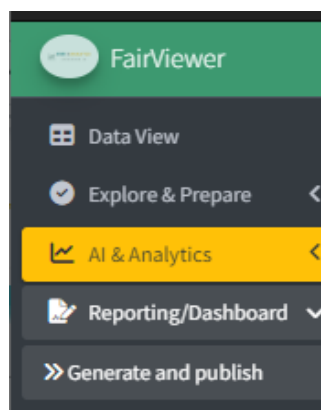


Fig 19 : Reporting Functionality

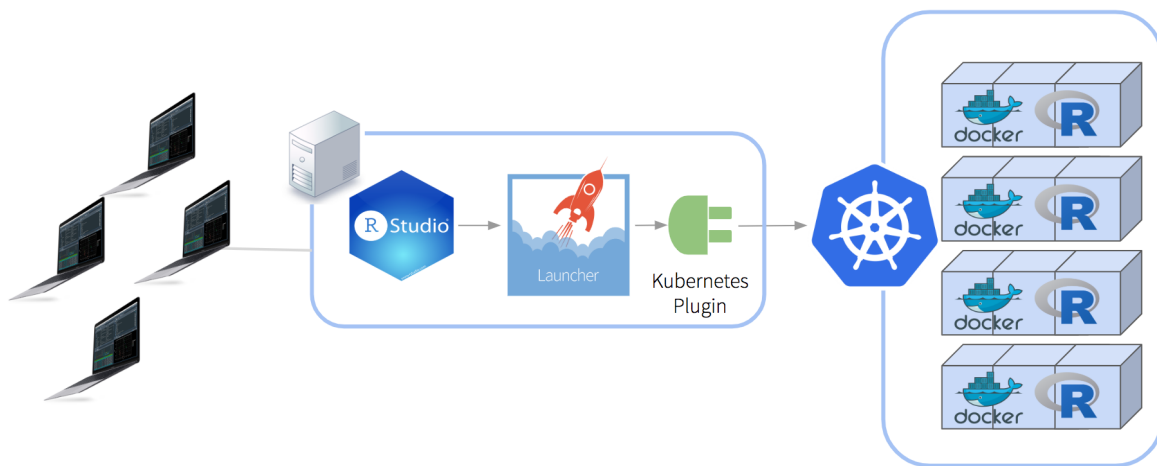
Next step, the user has to fill in a little information (Document Title, Author, Date, Document type). You have the ability to insert a comment. Multiple ways to generate and handle the reports are integrated into the app. Here are the different ways :

- Publish : the document can be published through a local network or into a personal database.
- Download : the report can be downloaded under multiple file formats (pdf, html, dashboard, presentation).
- Share : Sharing options by mail is considered by the app to facilitate documents' exchange.

Annex 1 : Application development & deployment.

The platform is developed using open source [R programming language](#) and based essentially on [shiny](#) packages for building the Web App. This technical choice makes our solution very stable and very interactive.

Our software is deployed using **Docker** on **kubernetes** hosted on AWS/AZURE/OVH. The illustration below gives an overview of our architecture to explain the deployment of our solution for every use.

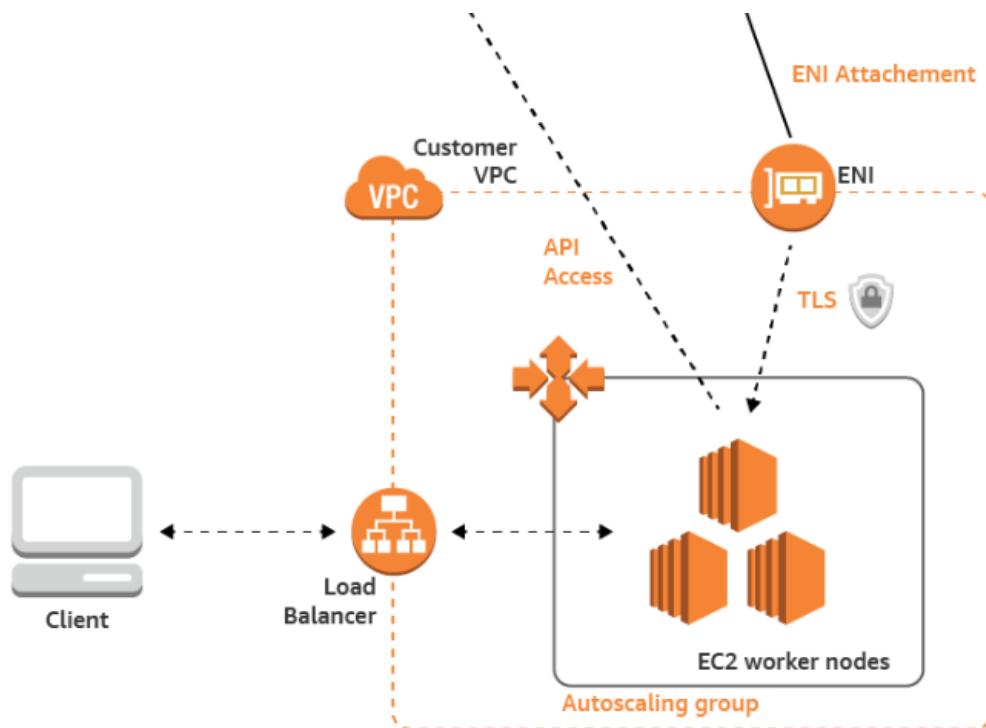


Annex 1 : Architecture of App's deployment

Annex 2 : GPRD COmpliance

For every use, here is the list of steps :

- New **virtual pod (Kubernetes)** is created from the **cluster**.
- The **communication** is internal and **virtual** inside the cluster.
- All the calculations are done on the fly **without** any intermediate **storage**.
- The pod is **destroyed** once the account is **logged out**.



Annex 2 : DataFlow Architecture

Annex 3 : IT Security overview

Our Solution has a very high security policy and many functionalities are integrated at all the stages to ensure complete protection.

As a default authentication and authorization process, we have implement on top of our platform [keycloak](https://www.keycloak.org/docs/latest/securing_apps/) with [OIDC provider](<https://openid.net/connect/>).

In addition, it can support various **authentications** (simple, IAM, LDAP,...).

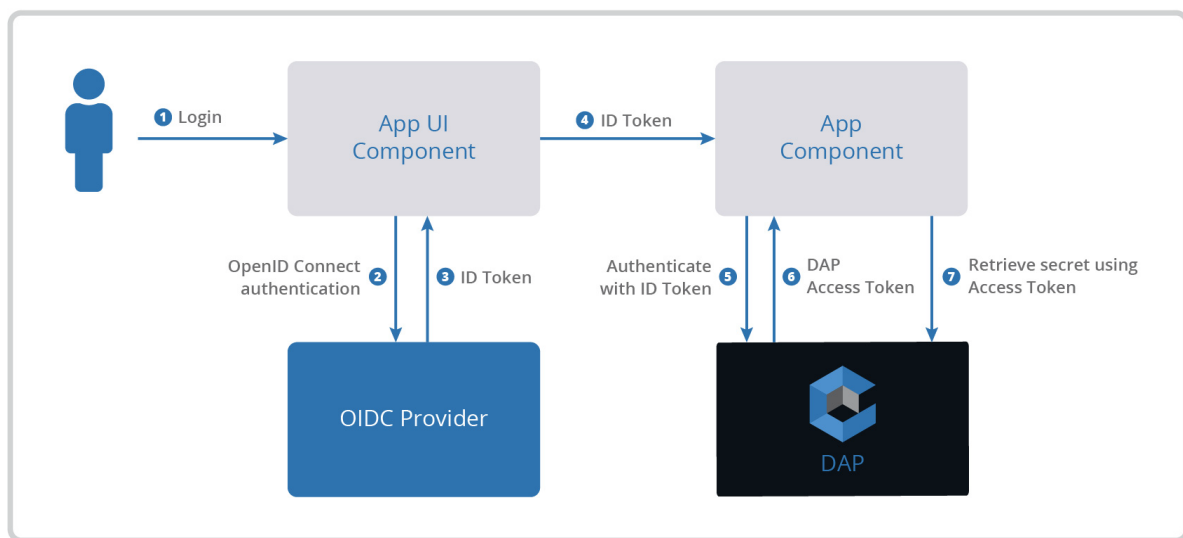


Fig : Keycloak based OIDC authentication protocol