

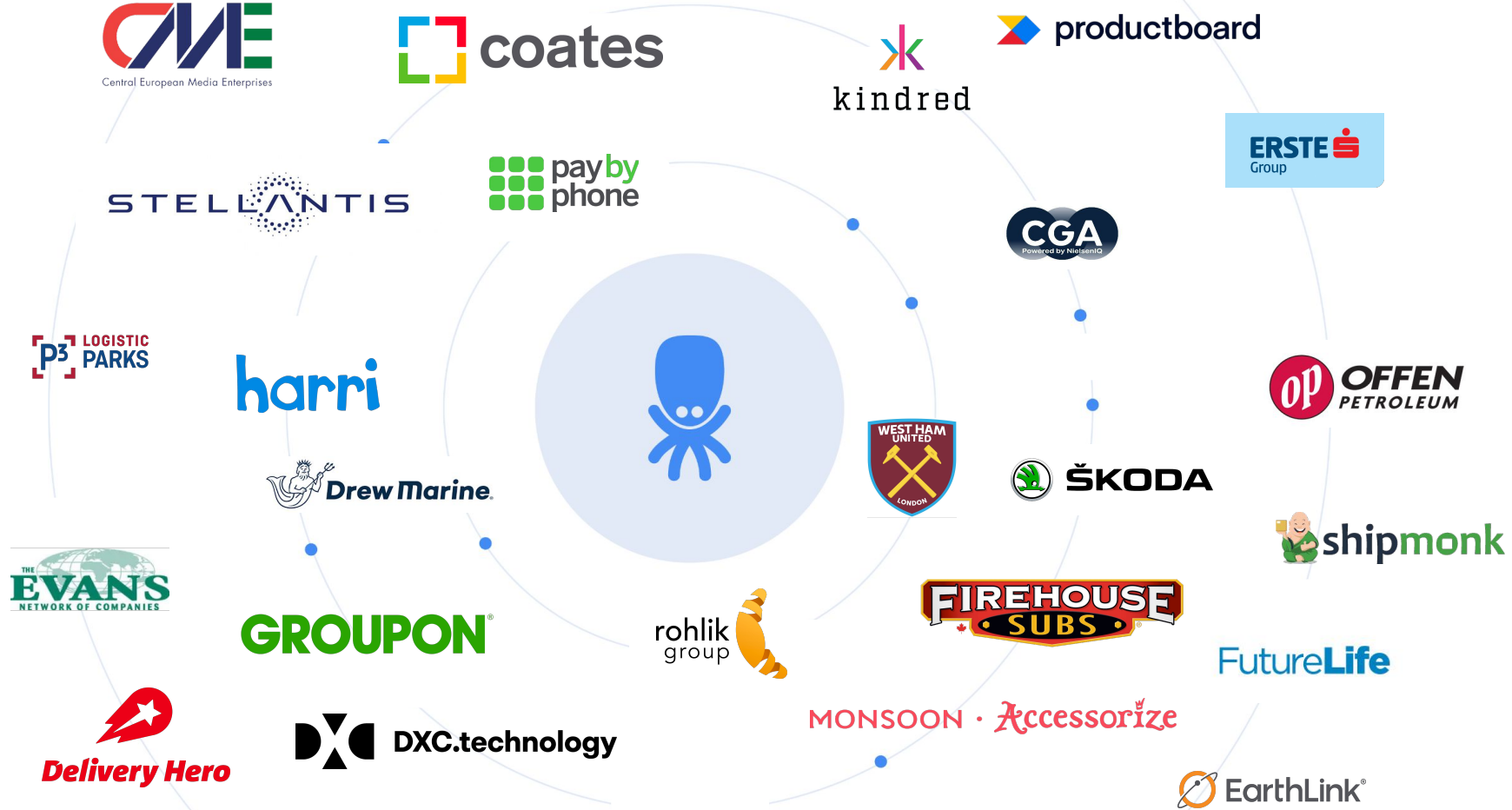
Modern Data Products **Powered by AI**

DATA

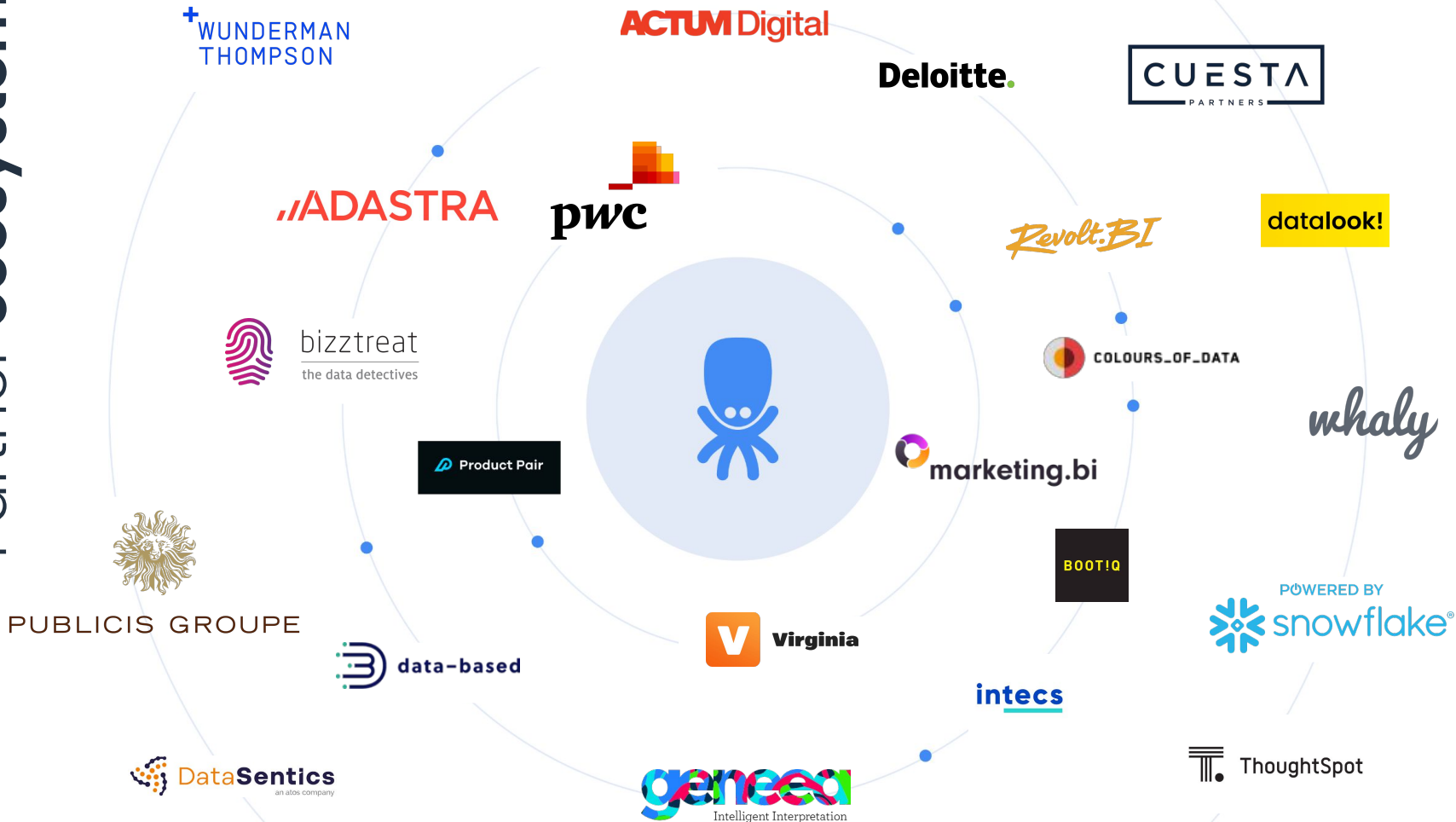
ACTIONS



250+ Client ecosystem



Partner ecosystem



Introduction

01

Our **History**

Where did it all start?



Founded in 2008 as a cloud migration consultancy

- Platform built to better serve clients and increase time to insights

Adopted the SaaS model in 2014

01/ Process layer

02/ For modern DWH

03/ From data → actions

04/ Empowering citizen users

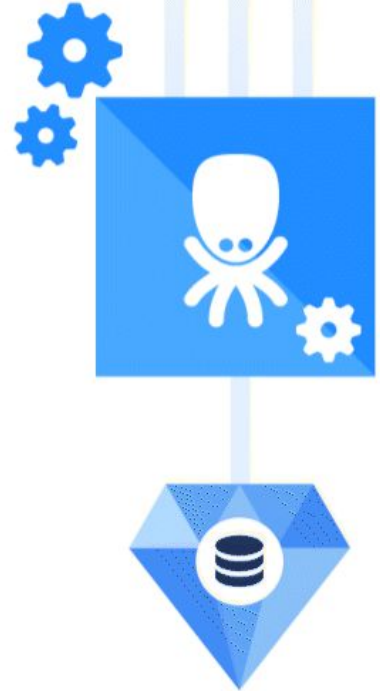
360TB/m, 3000 new
companies, \$100M new
revenue streams unlocked

Major retailer
\$100 million
In new projects in the 2 years

Olfen car grows
By 760%
Online revenue

FireHouse Subs support
1,200 franchises
With one person!

* Draft for information purposes only. All numbers are informational only and need to be verified.



Data Stack as a Service

02



DATA STACK AS A SERVICE

Keboola is a self-service, all-in-one data operations platform that simplifies and automates data operations, empowering business users to serve themselves whilst increasing operational efficiency.

Keboola's flexible and fully scalable platform is at the forefront of digital transformation, redefining the landscape by automating business processes and data management while enabling fast, fully governed data activation throughout the organization.

Focus on your data, not maintaining your infrastructure!

2014

Servicing all your data need since

270+

Customers worldwide

1000+

Users

zendesk

ERSTE

avast

lytics

KIWI.COM

makro

MALL.CZ

fourth

pricefx

GAP

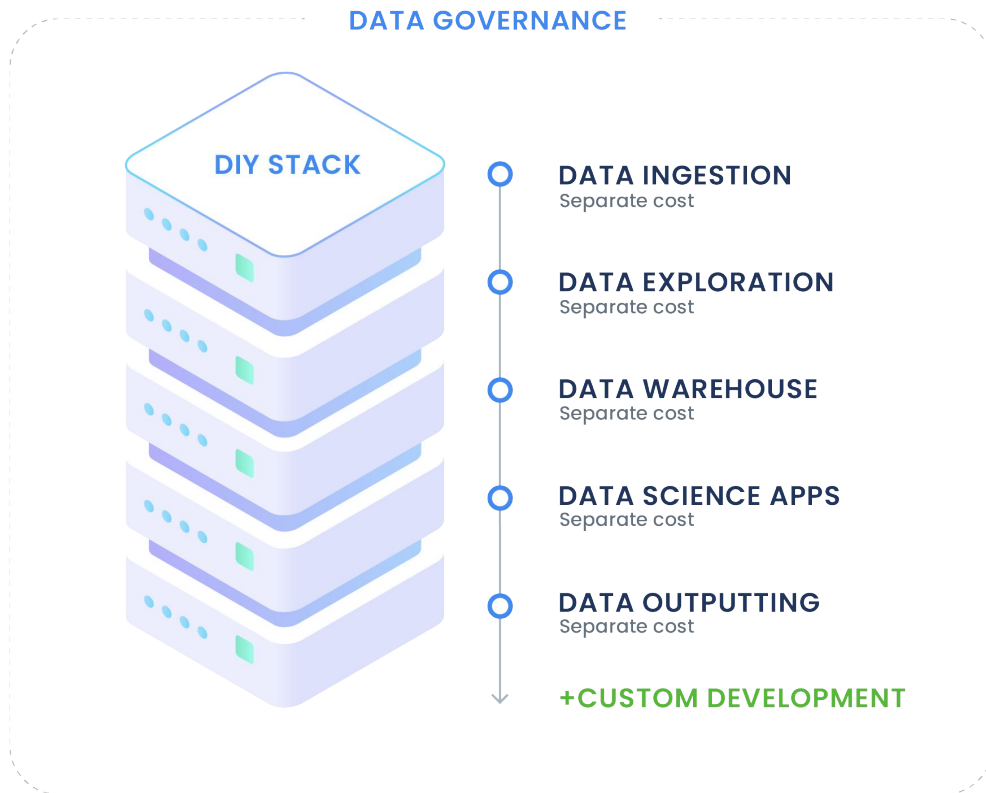
The Problem

Modern data stack?



Inherent challenges

- ✗ **Time-to-market** is still very slow
- ✗ **Scalability** is limited
- ✗ **TCO** is incredibly high
- ✗ **Maintenance** - Versioning, sharing, backups, archiving, tracking, access management, etc.



Our **Solution**

A managed, unified platform

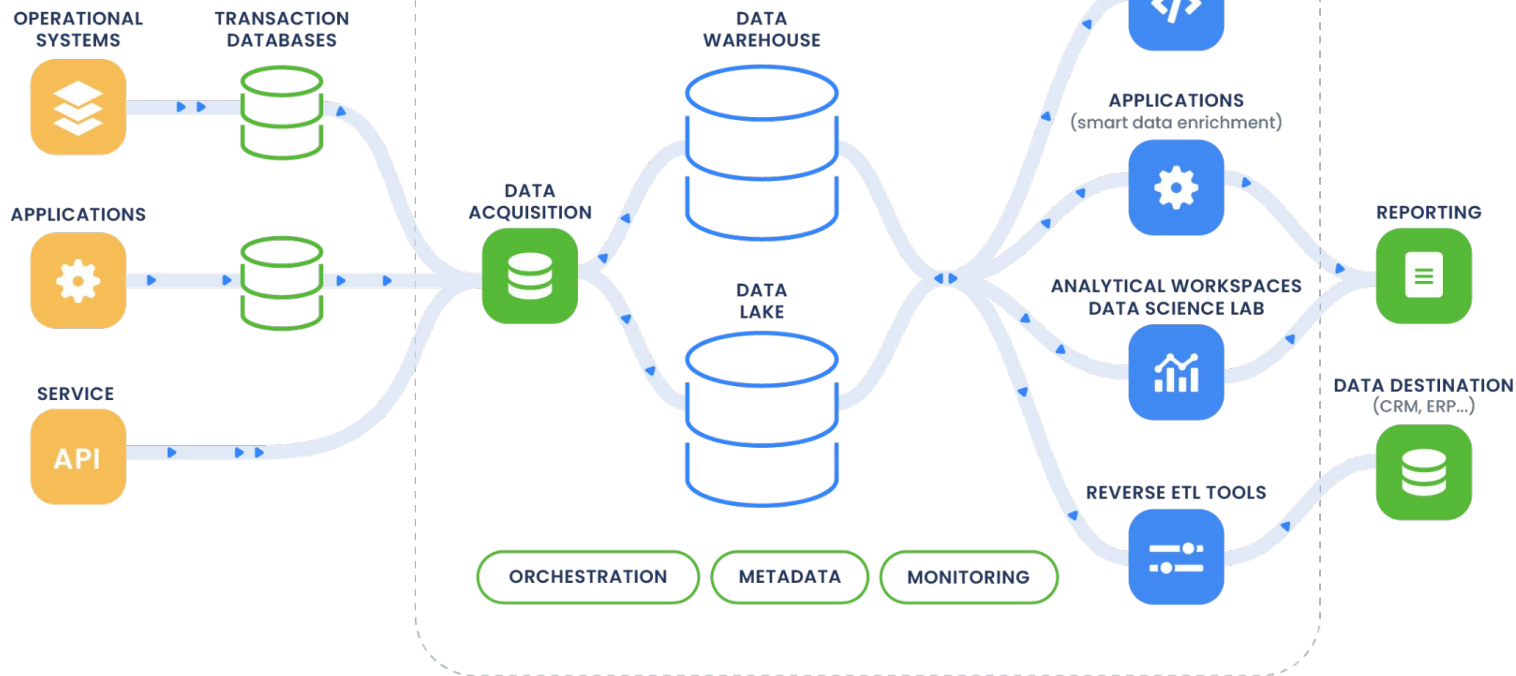
Inherent advantages

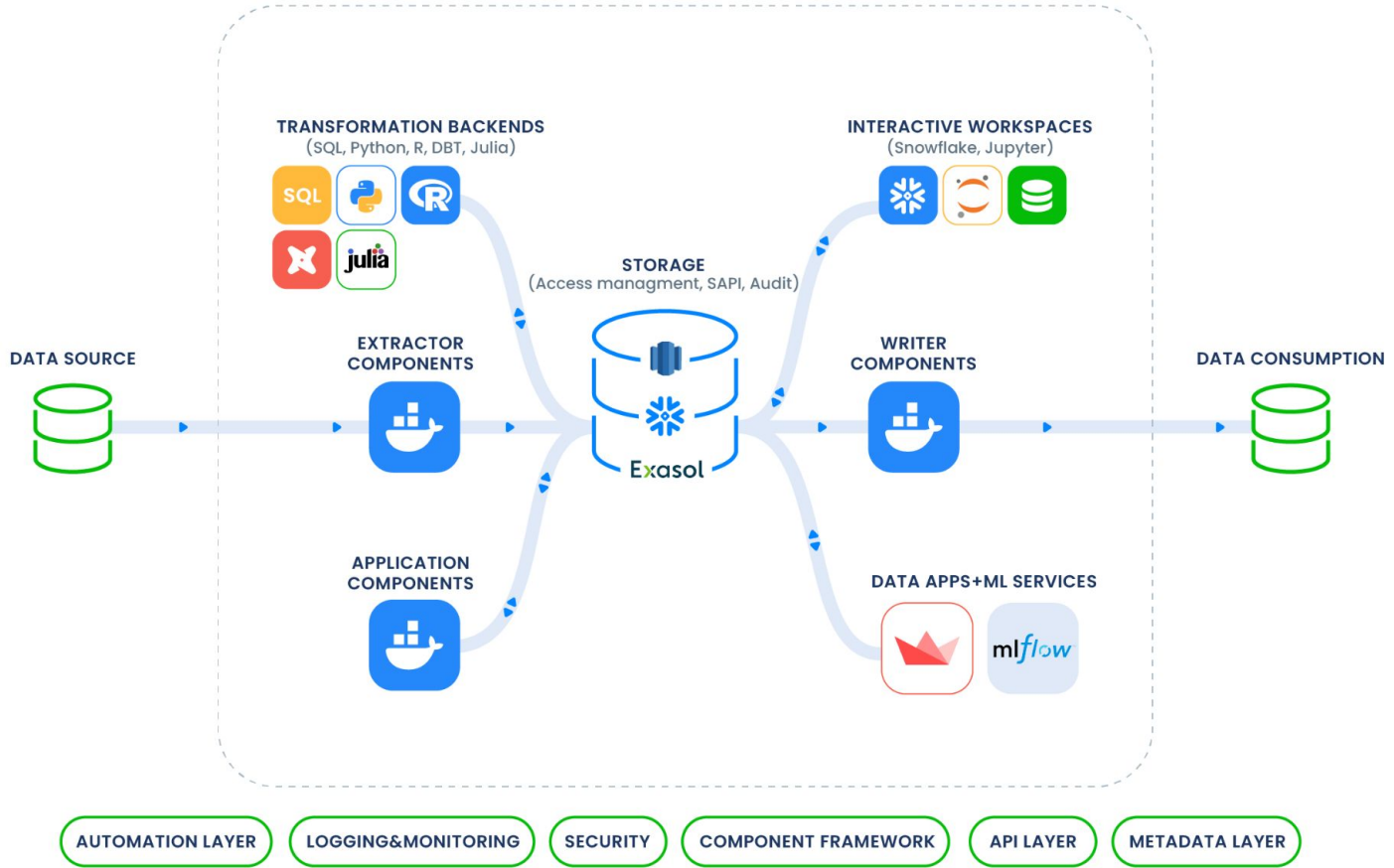
- ✓ **Time-to-market** is accelerated
- ✓ **Scalability** is seamless
- ✓ **TCO** is incredibly low
- ✓ **Maintenance** - that's on us!



One interconnected architecture that guarantees functionality of its building blocks while retaining its flexibility

Data Stack as a Service





General Overview

03

Keboola Project

- A Keboola Project is a logical construct creating an isolation of
 - Access rights
 - Responsibilities
 - Data pipelines
 - Storage
 - ... and processes embedded within it
- This isolation allows to distribute access and responsibilities among people and teams, but fully manage the data governance layer on the top of it.

The screenshot displays the Keboola Project Overview page. At the top right is the Keboola logo. The navigation bar includes 'Demo dbt demo', 'New Development Branch', a search bar, and icons for trash, notifications, and profile. The main navigation menu contains 'Dashboard', 'Flows', 'Components', 'Templates', 'Data Catalog', 'Storage', 'Transformations', 'Workspaces', and 'Jobs'.

Overview

ADD DESCRIPTION

Latest CLI Activity
5 days ago
Sync-pull command done. - [Settings](#)

Access
11 Users
5 API Tokens

Continue Your Work

- My Orchestration Trigger Queue V2 ...
Edited Apr 13 7:06 AM
- revolt_test
Edited Apr 4 4:22 PM
- Olist Flow
Edited Apr 4 4:18 PM
- Process Docs
Edited Mar 29 3:58 PM
- Olist Workshop - Keboola Storage
Edited Mar 14 3:17 PM
- Olist Workshop - External Snowflake
Edited Mar 14 3:16 PM

Project Users
MANAGE USERS

- Fisa Online
- David Pospisil Offline for 3 months
- Petr EU Šimeček Offline
- Václav Nosek Offline
- Milan Voverka Offline
- Roman Sigut Offline

Latest Activity
SEARCH ACTIVITY ALL DATA

Name	Duration / Version	Created / Last Change	Status
dbt Remote Snowflake: dbt beer demo - external infra (SFLK) Transformation	1 min 11 sec	34 minutes ago	Error
dbt Remote PostgreSQL: dbt Jaffle Shop demo - external infra (PostgreSQL) Transformation	1 min	34 minutes ago	Error
Snowflake: Send seed data to Snowflake - seed_breweries Data Destination	12 sec	35 minutes ago	Success
Snowflake: Send seed data to Snowflake - seed_beers Data Destination	13 sec	35 minutes ago	Success
PostgreSQL: PostgreSQL datastore - raw-seed-payments-csv Data Destination	20 sec	35 minutes ago	Success

Metadata Layer

- Every configuration can be further described with additional description.
- it brings the technical documentation closer to the code and may be used for automatic documentation and other integrations purposes.
- Transformation bucket, row.
- Any component configuration including Orchestrations (Flows).
- Storage: Buckets, tables, columns, column data-types, Data Catalogs.
- Project: Overall project description in the Overview tab, API tokens.



The screenshot shows the Keboola interface for a component named 'Send seed data to Snowflake'. The breadcrumb path is 'Components > Snowflake'. Below the title, there are tabs for 'Information & Settings', 'Notifications', and 'Versions'. The component is identified as a 'Data Destination' created by 'Keboola', used in '2 flows', with a link to 'Documentation'. A 'Data Destination Description' section is visible, containing a text area with the following content: '## Task' followed by 'Feeding data from seeds to remote DWH'. There are 'Generate Description with AI', 'CANCEL', and 'SAVE' buttons. A note at the bottom states 'Markdown is supported'.

The screenshot shows a 'Flow Description' dialog box. The text inside reads: 'This configuration defines a set of tasks to be run in a specific order. The tasks are:' followed by a bulleted list of three tasks: 'keboola.wr-google-sheets-19911362', 'keboola.dbt-transformation-23850126', and 'keboola.ex-google-drive-24157715'. Below this, it states 'The tasks are organized into three phases:' followed by another bulleted list: 'Data Extraction (phase 30602)', 'modelling stage (phase 64798)', and 'Data Push (phase 20014)'. The final paragraph explains the dependencies: 'The modelling stage phase depends on the Data Extraction phase, and the Data Push phase depends on the modelling stage phase. This means that the tasks in the Data Extraction phase must be completed before the tasks in the modelling stage phase can be run, and the tasks in the modelling stage phase must be completed before the tasks in the Data Push phase can be run.'

Logging

- Every job
- Every configuration change
- Every API call
- Every Storage Job, incl. sub-events and user previews
- Logging available for pre-set duration (configurable for dedicated deployments)
- Log accessible via API, Telemetry, Metadata Extractor in the form of datasets
- Activity Center



The screenshot displays the Keboola Jobs interface for a specific job. At the top, the navigation bar shows 'DEMO' and 'Production' environment. The main header indicates the job is 'SUCCESS'.

Configuration	Start	End	Created	Backend Size
Transformation: Input Tables Creation [IN-CRM-SALESFORCE] [9y2y4GgU] / Version 1	May 22 8:09	May 22 8:10	Scheduled run	Small

Key metrics and details:

- Duration:** +12 sec (Warning icon)
- Actual Duration:** 1 min 10 sec (Note: Looks like your job took longer than usual to finish)
- Parameters & Results:** Includes mode, component, config, and title.
- Mapping:** Lists input and output mappings with links to configuration pages.

A 'Duration & Errors' chart shows a series of bars representing previous runs, with a red dashed line indicating a threshold.

The 'Log' section provides a detailed event history:

Component	Event	Created
Snowflake SQL Transformation	Job '42247448' execution finished.	May 22 8:10
Snowflake SQL Transformation	Output mapping done.	May 22 8:10
Storage	Deleted workspace snowflake:WORKSPACE_42247510	May 22 8:10
Storage	Fetches details of table out.c-crm-tr-9y2y4GgU.opportunity_stage	May 22 8:10
Storage	Added keboola.snowflake-transformation metadata to table out.c-crm-tr-9y2y4GgU.opportunity_stage	May 22 8:10

Observability

“Trio” of Ops components:

- Keboola Telemetry
- Project Metadata
- Query History

Usually orchestrated in dedicated “Ops” project

Add-on called Keboola Activity Center – 360 observability in BI tool



The screenshot displays the Keboola Activity Center interface, showing three data source configurations. Each configuration is presented in a card-like view with a header, a table of configurations, and an 'ADD NEW CONFIGURATION' button.

KBC Project Metadata

Data Source

<input type="checkbox"/> Name	Data	Used In	↑ Last Change	⇅ Last Use
+ ADD NEW CONFIGURATION				
<input type="checkbox"/> Organization Metadata	Show Data	-	May 3 12:10 PM Fisa	May 3 12:10 PM Fisa ✓

Telemetry

Data Source

<input type="checkbox"/> Name	Data	Used In	↑ Last Change	⇅ Last Use
+ ADD NEW CONFIGURATION				
<input type="checkbox"/> BYODB Telemetry	-	-	May 3 11:52 AM Fisa	No run recently

Snowflake Query History

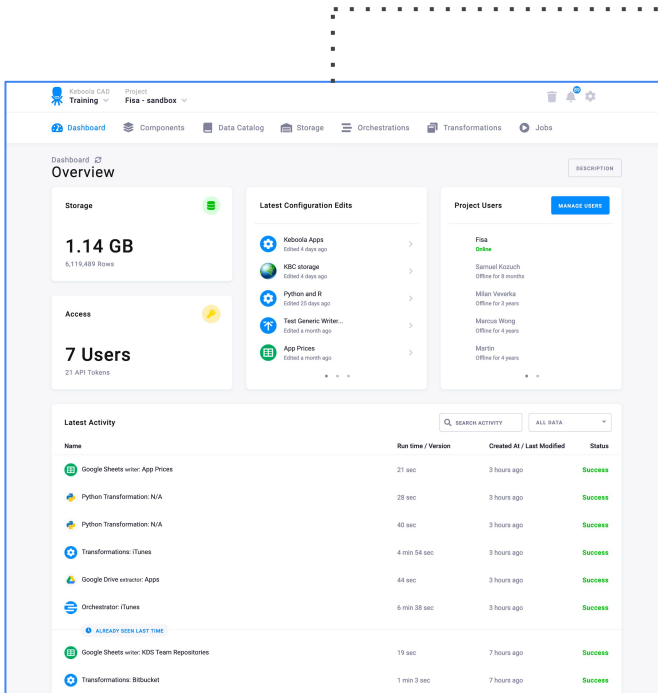
Data Source

<input type="checkbox"/> Name	Data	Used In	↑ Last Change	⇅ Last Use
+ ADD NEW CONFIGURATION				
<input type="checkbox"/> BYODB query history	Show Data	1 flow	Apr 25 1:39 PM Fisa	37 minutes ago Scheduled run ✓

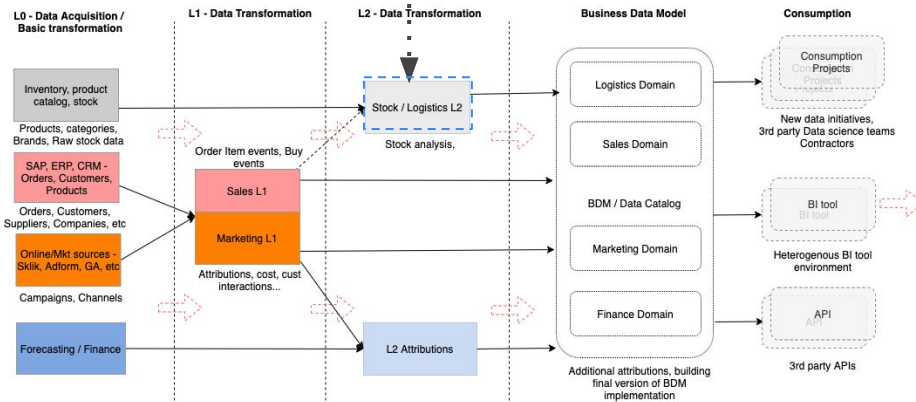
Feature Spotlight

04

Multi-Project Architecture (MPA)

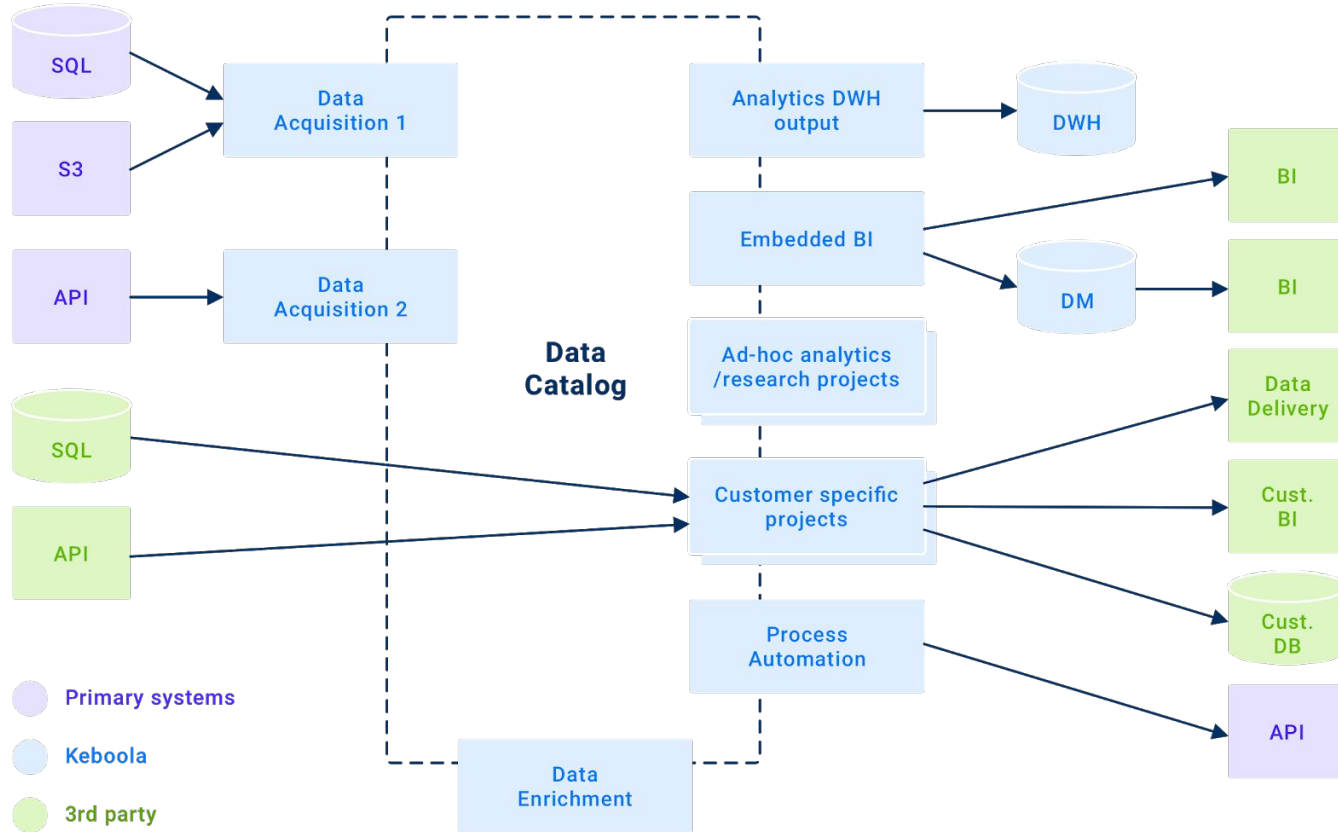


Single Keboola project



Keboola Enterprise Multi-Project Architecture (MPA)

Data **Tool** -> Data **Platform**

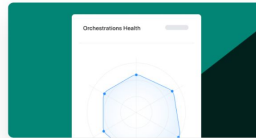


Activity Center

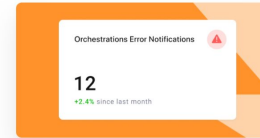
Complete observability of your environment



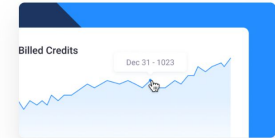
- Observability layer
 - Cost attribution
 - User 360 & Project 360
- FinOps: What/who costs how much
- SecOps: Observe 80+ different actions
- Quality Ops: Quality of the pipeline, orchestrations, credits optimization



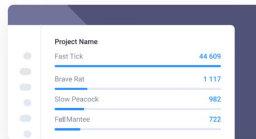
Orchestrations Health Spider



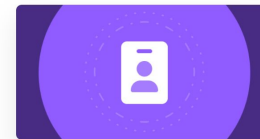
Health of your workspace



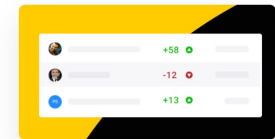
Billed Credits



Identify Financial Outliers



Security events



Biggest Gainers, Biggest Droppers

Activity Center

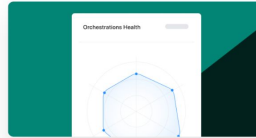
Complete observability of your environment



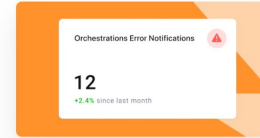
- Custom integration of **metadata** and **telemetry data** – proprietary data model

- **BI Templates**

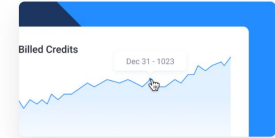
- Tableau
- PowerBI
- Looker
- ThoughtSpot



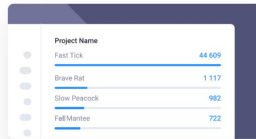
Orchestrations Health Spider



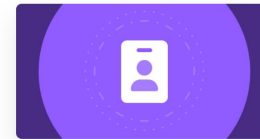
Health of your workspace



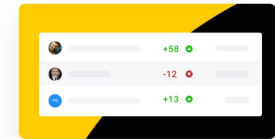
Billed Credits



Identify Financial Outliers



Security events



Biggest Gainers, Biggest Droppers

No-code Transformations

Empowering citizen users



- Build data products without coding or IT
- Clean data in a couple of clicks
- Collaborate with coders in the same environment
- Scale - Automate and reuse your no-code transformations

A screenshot of the Keboola Transformations interface. The interface is titled "Transformations Demo" and includes a "RUN TRANSFORMATION" button in the top right corner. Below the title bar, there are tabs for "Information & Settings", "Notifications", and "Versions". The main content area is divided into three sections: "Data", "Data Manipulations", and "Active Data Manipulations". The "Data" section shows a table named "IN GOOGLE ADS" with a "campaigns" column. The "Data Manipulations" section has tabs for "Clean & Format", "Create & Update Data", "Filter Data", "Do Math", and "Combine". Under "Clean & Format", there are six manipulation cards: "Replace Data", "Rename Column", "Remove Duplicate Rows", "Remove Column(s)", "Clean Data", and "Format Date & Time". The "Active Data Manipulations" section shows three active manipulations: "Remove Column(s) customerid", "Rename Column name, Name", and "Filter Rows servingStatus = ENDED".

Data Templates

Pre-built use cases

- Deploy end-to-end data pipelines in minutes
- Business users can self-serve their IT needs
- Modify existing templates or build your own
- Supported in CLI



Template Library



Add New Template

Categories

- Analytics 5
- CRM 9
- Data Governance 1
- E-commerce 6
- Git 4
- Marketing 4
- Project Management 4
- Surveys 4
- Tutorials 3

All Templates (40)

Sort A to Z



Advertising Platforms to BigQuery

Aggregate your advertising data to optimize the monetization of your online traffic and significantly increas...

+ USE TEMPLATE



Advertising Platforms to Google Sheets

Aggregate your advertising data to optimize the monetization of your online traffic and significantly increas...

+ USE TEMPLATE



Advertising Platforms to Keboola DWH

Don't you have your own data warehouse? Use Keboola provided Snowflake database. Aggregate your advertising...

+ USE TEMPLATE

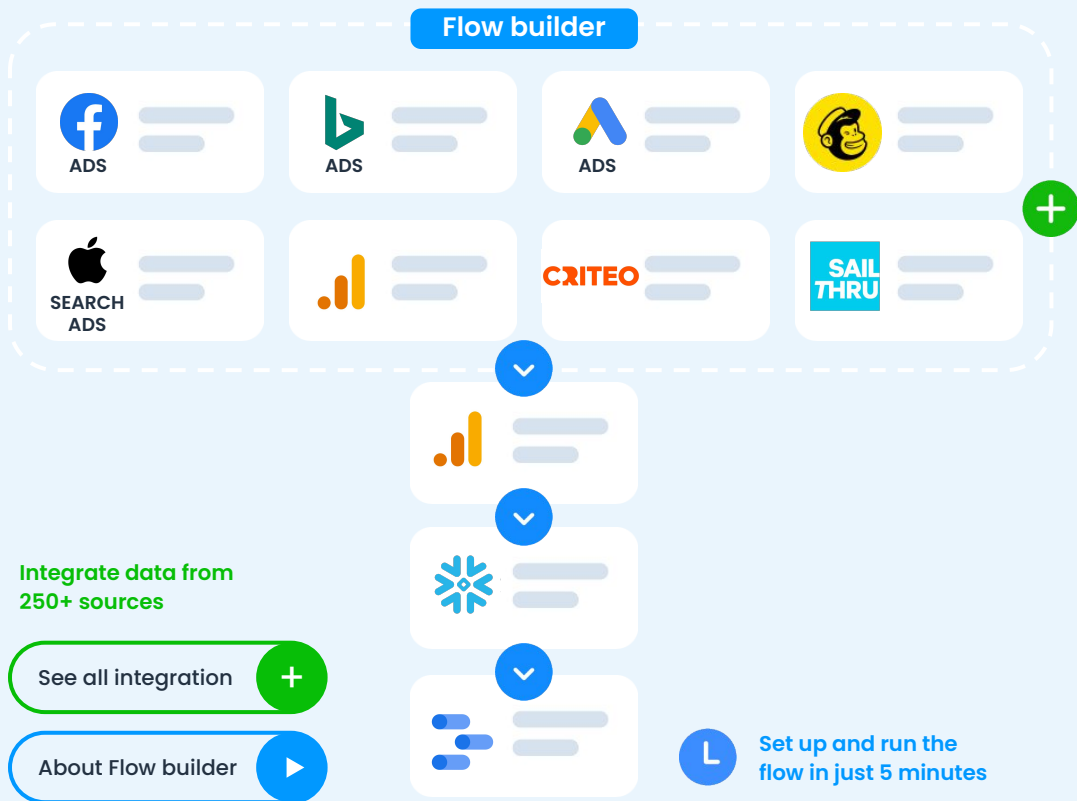


Advertising Platforms to Snowflake

Aggregate your advertising data to optimize the monetization of your online traffic and significantly increas...

+ USE TEMPLATE

Marketing analysis



Highlights box:

Campaign analytics is bases for demand forecast, dynamic pricing, CDXP and others.

Usually can be run within two-three weeks from ground up

Benefit box:

Based on Keboola standardize model enables fast time to market with big flexibility to add proprietary sources.

Whole data product unlocks fast way to profit driven marketing

Use cases

05

Turn your Data **into AI Solutions**

Our recipe for success

- Build sales and financial forecasts
- Optimize your inventory and delivery routes
- Personalize marketing communications at scale
- Turn your machine-learning algorithms into easy-to-consume data products

But this is just the beginning...



Olfin car
760% increase in sales
in a single quarter



CSC - Marketing Analytics

#Marketing #ROI



Background

- Provider of IT services, professional services and solutions for large corporations
- \$12B in annual revenues
- Roughly 74,000 employees worldwide
- Over 3,500 sales staff and 500 person marketing department.
- mid-8 digit marketing budget annually

Challenges

- 1) How are our marketing efforts affecting our sales?
- 2) How are our digital properties performing?
- 3) Can we show direct and indirect links between marketing spend and resulting sales?
- 4) Difficult to identify where to start

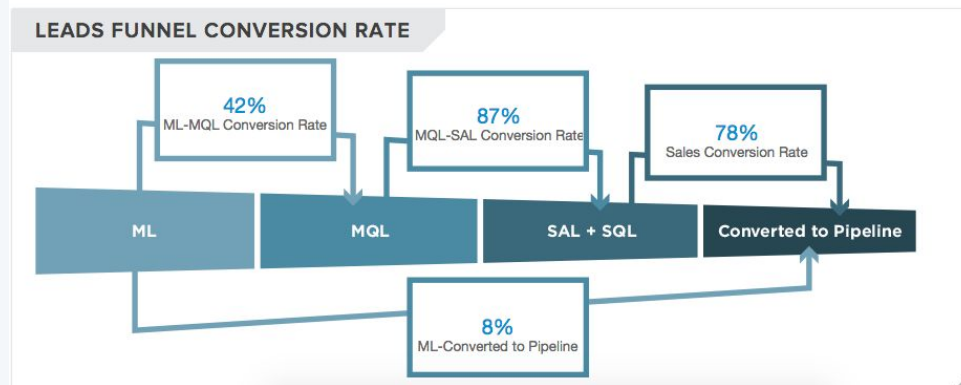
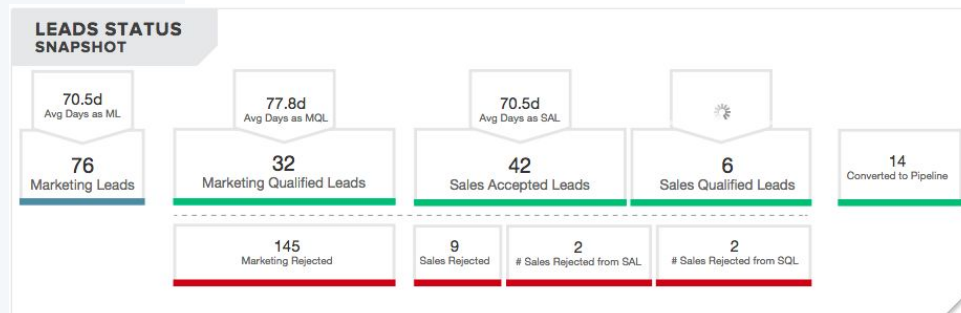


CSC - Marketing Analytics

#Marketing #ROI



- Everything started with Salesforce
- Integrating 50+ sources into CRM (including social media and offline)
- Give executives:
 1. Effect of marketing on bottom line
 2. End to end solution
 3. Connect CRM to BI tools for more complex analysis and reporting when needed

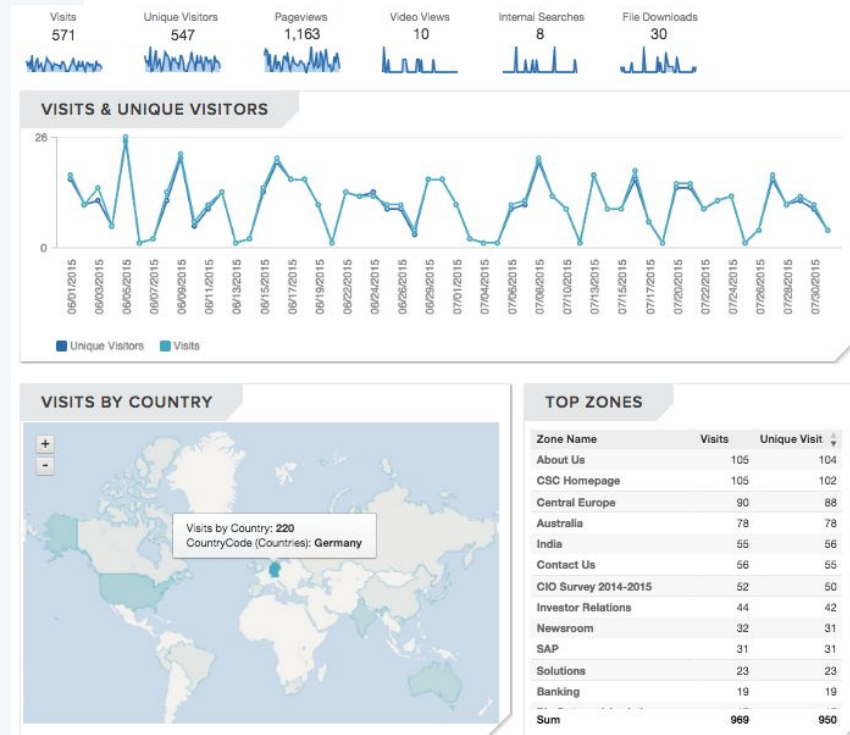


CSC - Marketing Analytics: The result

#Marketing #ROI



- 90 days to launch fully working solution
- Demonstrate 206% growth in MQL
- Instant access to insight
- All digital data into a single platform



Attribution data model use-case

Conversion Path

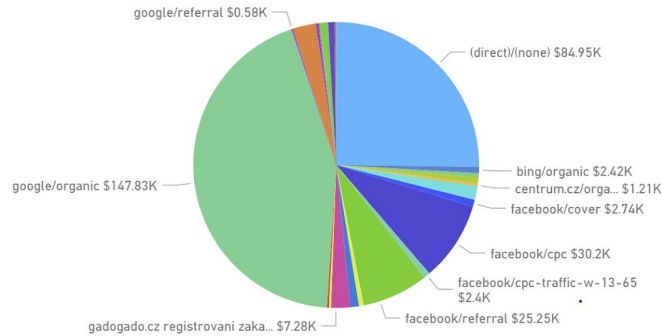
Attribution Model Type

first click
 linear
 last click
 LND click

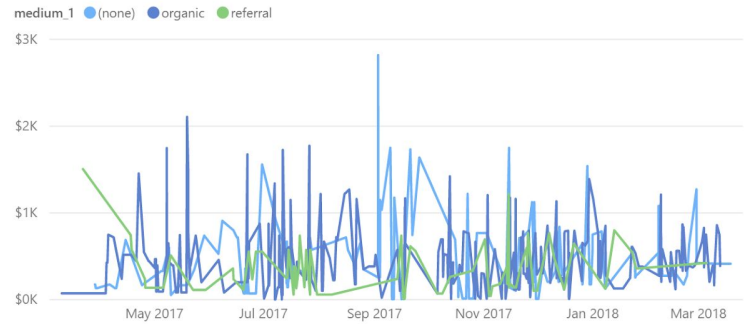
Attribution Lookup Window

1
 14
 90
 7
 30

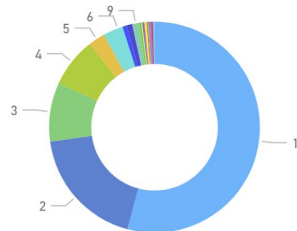
\$ Revenue by Attribution Model



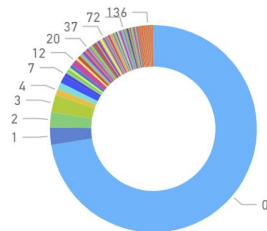
Top 3 Mediums by Revenue



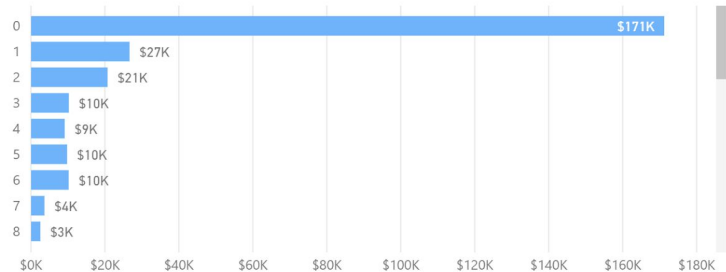
Visits to Order



Days to Order



\$ Revenue by Path Length (# Days)



Unibet - Predictive Data-driven Marketing

1. Challenge

Unibet is a part of Kindred Group - an online gambling operator - and provides a platform for sports betting, casino, games and online poker. Today, Unibet has 9.5 million customers in over 100 countries and over 650 employees.

Troubles:

- Struggling with effective marketing planning caused by fragmented data and poor attribution across different marketing channels
- Campaign evaluation proved difficult due to disparate marketing sources across multiple regions
- Inability to effectively allocate its marketing budget within different channels

2. Big Goal

Introduce a platform to enable marketing teams to effectively allocate budgets based on predictive analytics

3. Data sources

- AdWords, Facebook Ads, Twitter Ads, Bing Ads, Adform, AppNexus and others
 - Each of them having dozens location mutations

4. Solution

- Keboola Connection stands as a single platform for the entire marketing data sources integration process.
- On the top of the data, spend and churn prediction models are computed.
- Everything is visualized in Qlik and PowerBI.

5. Result

Ability to effectively predict how much to spend per channel and plan the marketing campaigns as well as overall marketing strategy based on data, not feelings.



Keboola RTB optimization - Axa & AdExpress

1. Challenge

Our client from the financial sector spends high amounts on performance campaigns.

Marketing department didn't have **clarity/attribution** of RTB campaigns in the whole funnel to the last click **conversion**.

- multiple disconnected data sources
- local market specific data sources
- **1.5 TB** of data to be analyzed daily using sophisticated ML algorithms

2. Big Goal

Calculate attribution of RTB **re-marketing** on the last click.

Ability to identify most **profitable** acquisition combinations in the funnel.

3. Data sources

- **Google Analytics**
- **RTB Adform**
- **Spreadsheets**

4. Solution

Data is integrated in KBC ETL, where data is cleaned, connected and prepared for further analysis.

Advanced transformation and ML predictions are done in **Python**. The entire advertising funnel is mapped and attributions calculated using multiple algorithms.

Result:

2x more conversions

2.5x higher conversion value

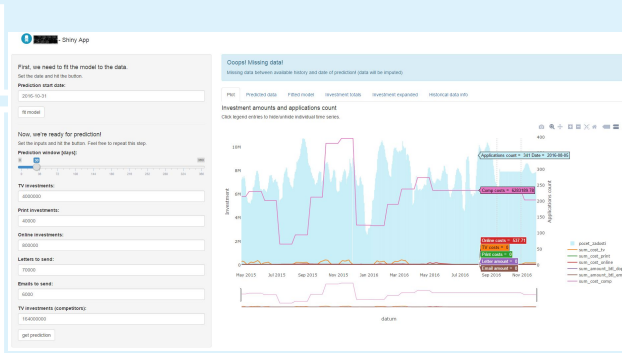
#marketing spend prediction

Prediction app for marketing spends by channel

1. Challenge

A marketing department at a large financial institution needs to **plan and predict** spend on each channel with **different** goals in mind.

- multiple disconnected data sources
- long funnel customer journey
- hard to **predict** the future with ML in user **friendly** environment



3. Solution

aLook Data Science agency has built a product on top of Keboola which predicts spend and marketers can use it simple by adjusting weights in graphical interface.

Since then, aLook has resold this product many times.

3. Data sources

- **Google Analytics**
- **RTB Adform**
- **Spreadsheets**

Keboola + Geneea - Airlines social media text analysis

Keboola Connection was used to extract Facebook data, the Geneea NLP application within Keboola Connection was used to determine sentiment and produce the metadata for each post.

Promatix used Tableau to visualise the data.

Click on the logos to the right to find out more about the companies and technologies involved in this work.

Show Top ... Airlines by Volume of Posts

50



178,405

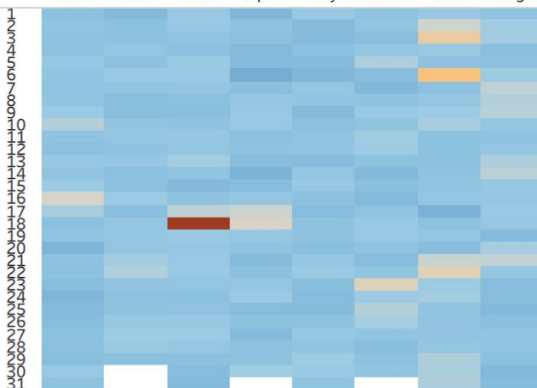
Posts Analysed in this Dashboard



Powered by + a | b | e | a | u

Sentiment Calendar

(Click to filter dashboard to single date)



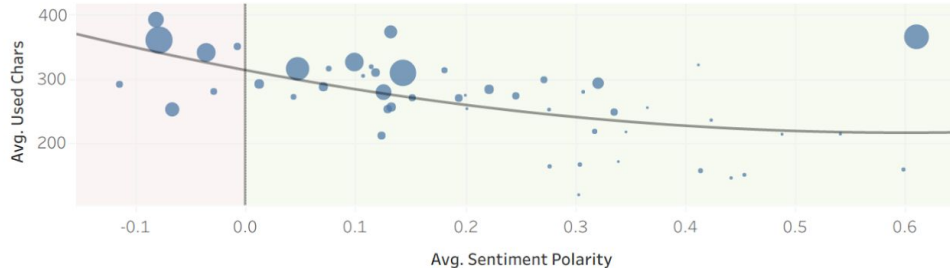
Volume of Posts and Average Sentiment by Airline

(Hover to view detail of posts, click to fix the dashboard to this airline)



Sentiment vs Length of Post

(hover over to filter/highlight airline)



Whats is Being Posted

(Select type of content in drop down)

tag

tag	count
airlines	63,409
flying	32,153
American Airlin..	23,900
Southwest Airli..	22,964
United	20,335
Delta Air Lines	18,858
plane	17,416
airport	16,550
delay	15,737
Australia	14,913
Qantas	11,654
easyJet	11,263
airways	11,252
bags	9,395
flight	8,405
JetBlue Airways	8,398
Jetstar	8,180
family	7,622
luggage	7,373
virgin atlantic	6,419

Annex

06

Keboola Enables you to Supercharge your Business with AI

Artificial Intelligence (AI) integration

✓ Available now in beta

- DataOps automation with AI
- Every data person's robotic friend

🚀 Coming Soon

- A **text-to-pipeline solution** that creates new data pipelines from simple text instructions
- **Custom-trained AI models** for tailored applications
- Automated impact analyses
- Complete AI-enablement solution to quickly and painlessly integrate with your existing data stack



Keboola Automates DataOps with AI

Automate your DataOps with AI



- Error messages are readable by humans
- Automated data governance

A screenshot of a web browser window displaying an error message. The window has a dark blue header with three colored window control buttons (red, yellow, green) on the left. The main content area is white and contains a red circular icon with a white exclamation mark. Below the icon, the text reads "Job Error" followed by a long alphanumeric string and a close button (X). The error message itself is a JSON object: {"error": "Invalid request", "exceptionId": "com-keboola-azure-north-europe-connection-ba4f705c536a6660904cb028513524cb", "errors": [{"key": "primaryKeysNames", "message": "Primary keys on columns [order_id,row_nr] cannot be created. Primary keys columns must be set nullable false."}], "code": "validation.failed"}. Below the error message is a grey box with an "AI Explanation" section. The explanation text states: "The error is caused by an attempt to create a table with primary keys on columns `order_id` and `row_nr` that are set to `nullable true`. To fix the issue, the `nullable` property of the columns should be set to `false`." At the bottom of the window, there are two buttons: a white button with a refresh icon and the text "RELOAD EXPLANATION", and a green button with a speech bubble icon and the text "CONTACT SUPPORT".

AI for Descriptions

Transformations
10newest_popular_shows

Information & Settings Notifications Versions

Type Transformation Used in 1 flow Data Show Data Documentation Licence

Transformation Description

+ Add Description or Generate Description with AI

Automatically generates a description using advanced artificial intelligence algorithms. The generated description may not be perfect, but it can serve as a starting point for further editing.

Transformations
10newest_popular_shows

Information & Settings Notifications Versions

Type Transformation Used in 1 flow Data Show Data Additional Links Documentation Licence

Transformation Description

This configuration creates a table called `ten_newest_shows` that contains the title, director, listed_in, description, and date added of the 10 newest TV shows from the `netflix` table. The date is converted to a date format using the `try_to_date` function.