

夏休みの自由研究

～ OCIで始めるデータ分析 ～

2021/7/20 OCIjp#20

自己紹介

白濱 亮（しらはまりょう）

- ・ 所属
 - ・ インフラエンジニア
- ・ OCI経験
 - ・ 2年くらい
 - ・ 普段はIaaSを使った環境構築が多いですが最近ではPaaSにも手を出しています
- ・ 好きなOCIサービス
 - ・ Analytics
 - ・ Functions



今回の発表で伝えたいこと

- ✓ OCIサービス中のAnalyticsの位置づけが何となく分かること
- ✓ Analyticsを使うと何ができるのか何となく分かること
- ✓ 業務の中でAnalyticsの使い所が何となくイメージできること

OCIサービス一覧

Compute

Artifact Registry
Compute
Container Engine for Kubernetes
Container Registry
OS Management
Oracle Cloud VMware Solution
Oracle Linux

Storage

Archive Storage
Block Volume
Data Transfer
File Storage
Object Storage
Storage Gateway

Governance & Administration

Audit
Budgets
Cloud Advisor
Cost Analysis
Support Management
Tagging

Integration

Apiary
Integration

Networking, Edge, and Connectivity

DNS and Traffic Management
Email Delivery
FastConnect
Health Checks
Load Balancing
Network Load Balancer
Networking

Application Development

API Gateway
Application Express (APEX)
APEX Service
Blockchain Platform
Content and Experience
DevOps
Digital Assistant
Events
Functions
Mobile Hub
Visual Builder
Visual Builder Studio
WebLogic Server for OCI
WebLogic Server for OKE

Analytics and AI

Analytics Cloud
Fusion Analytics Warehouse
Language
Streaming

Data Management

Autonomous Data Warehouse
Autonomous Transaction Processing
Bare Metal and Virtual Machine DB Systems
Big Data
Data Catalog
Data Integration
Data Flow
Data Safe
Data Science
Database Migration
Exadata Cloud Service
Exadata Cloud@Customer
External Database
GoldenGate
MySQL Database
NoSQL Database
Service Connector Hub
Hybrid Cloud
Roving Edge Infrastructure

Observability and Management

Application Performance Monitoring
Database Management
Java Management
Logging Analytics
Logging
Management Agent
Monitoring
Notifications
Operations Insights
Oracle Management Cloud
Resource Manager

Security, Identity, and Compliance

Security Guide
Bastion
CASB
Cloud Guard
Compliance Documents
Identity and Access Management
Security Advisor
Security Zones
Vault
Vulnerability Scanning
Web Application Firewall

Marketplace

Cloud Infrastructure Marketplace
Cloud Marketplace

OCIサービス一覧

Compute

Artifact Registry
Compute
Container Engine for Kubernetes
Container Registry
OS Management
Oracle Cloud VMware Solution
Oracle Linux

Storage

Archive Storage
Block Volume
Data Transfer
File Storage
Object Storage
Storage Gateway

Governance & Administration

Audit
Budgets
Cloud Advisor
Cost Analysis
Support Management
Tagging

Integration

Apiary
Integration

Networking, Edge, and Connectivity

DNS and Traffic Management
Email Delivery
FastConnect
Health Checks
Load Balancing
Network Load Balancer
Networking

Application Development

API Gateway
Application Express (APEX)
APEX Service
Blockchain Platform
Content and Experience
DevOps
Digital Assistant
Events
Functions
Mobile Hub
Visual Builder
Visual Builder Studio
WebLogic Server for OCI
WebLogic Server for OKE

Analytics and AI

Analytics Cloud
Fusion Analytics Warehouse
Language **データ分析・AI**
Streaming

Data Management

Autonomous Data Warehouse
Autonomous Transaction Processing
Bare Metal and Virtual Machine DB Systems
Big Data
Data Catalog
Data Integration
Data Flow
Data Safe
Data Science
Database Migration
Exadata Cloud Service
Exadata Cloud@Customer **データ基盤**
External Database
GoldenGate
MySQL Database
NoSQL Database
Service Connector Hub
Hybrid Cloud
Roving Edge Infrastructure

Observability and Management

Application Performance Monitoring
Database Management
Java Management
Logging Analytics
Logging
Management Agent
Monitoring
Notifications
Operations Insights
Oracle Management Cloud
Resource Manager

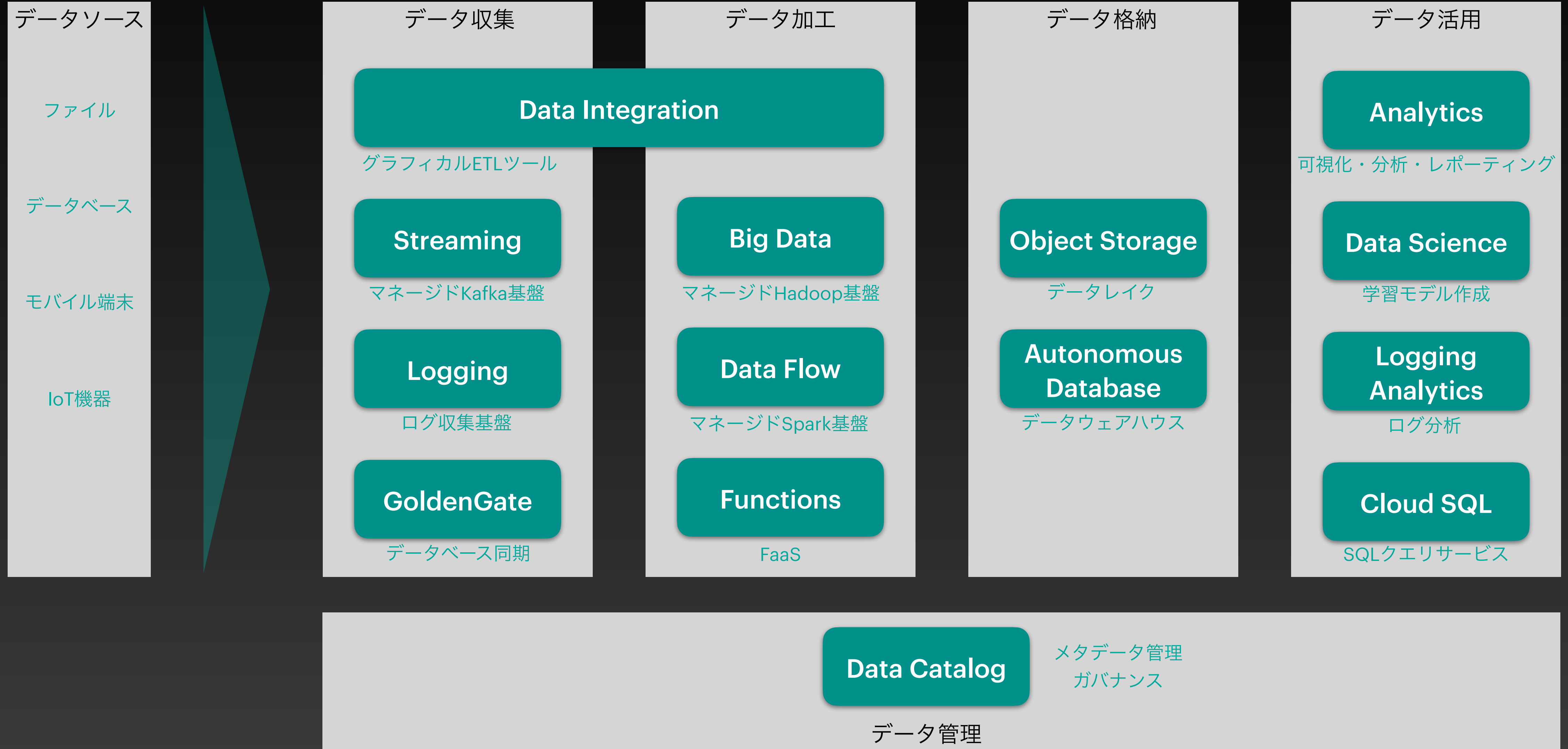
Security, Identity, and Compliance

Security Guide
Bastion
CASB
Cloud Guard
Compliance Documents
Identity and Access Management
Security Advisor
Security Zones
Vault
Vulnerability Scanning
Web Application Firewall

Marketplace

Cloud Infrastructure Marketplace
Cloud Marketplace

データ関連サービス



Analytics Cloudとは

高機能なBIサービス



セルフサービス Analytics

ユーザーフレンドリーな直感的な操作性で
業務担当者による
データの準備・加工や可視化・分析を強力に支援

エンタープライズ Analytics

全社共通の分析モデルの作成と情報展開の実現
ビジネスの概況を直感的に把握できる「ダッシュボード」
自動レポートイングおよび配信機能など

拡張分析

データの特徴を自動診断・解説
ユーザーによるインサイトの獲得を強力に支援

Analyticsの利用にはIDCSが必須

SaaS

PaaS

Analytics
Contents & Experience
etc...

IaaS

Compute
Storage
Network
etc...

IAMユーザで利用可能

IDCSによるID管理が必要

Analyticsの価格体系

価格体系	単位	プロフェッショナル	エンタープライズ
OCPU課金	OCPU per hour	¥118	¥236
ユーザ数課金	User per month	¥1,760	¥8,800

※ \$1=¥110で計算

※ ユーザ数は10-3000人で設定可能。スペックは4 OCPU相当。

他社製品

クラウド統合



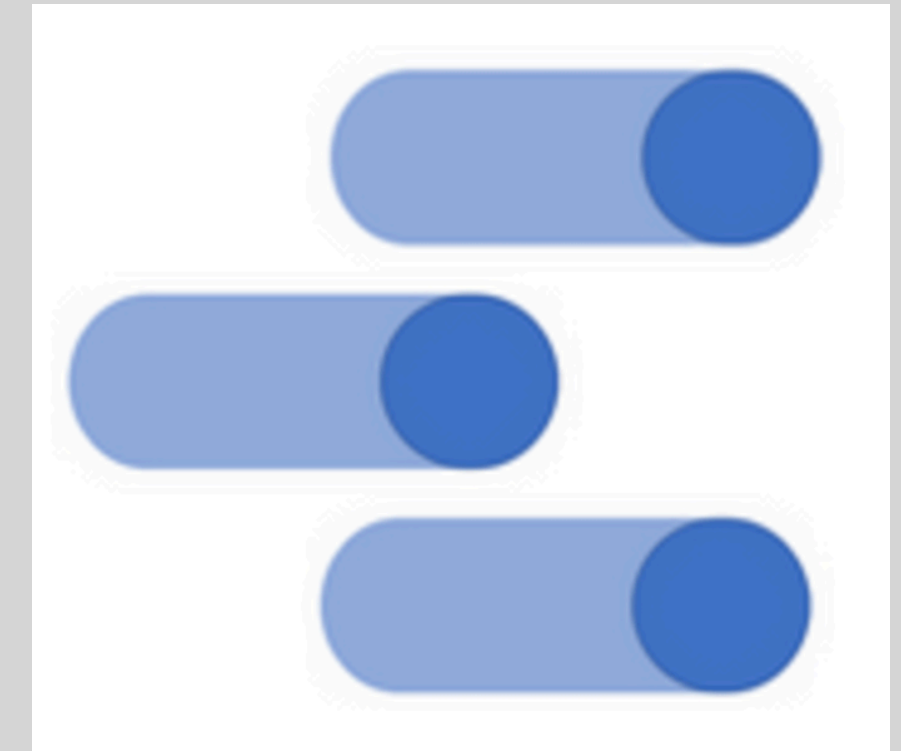
Analytics



QuickSight

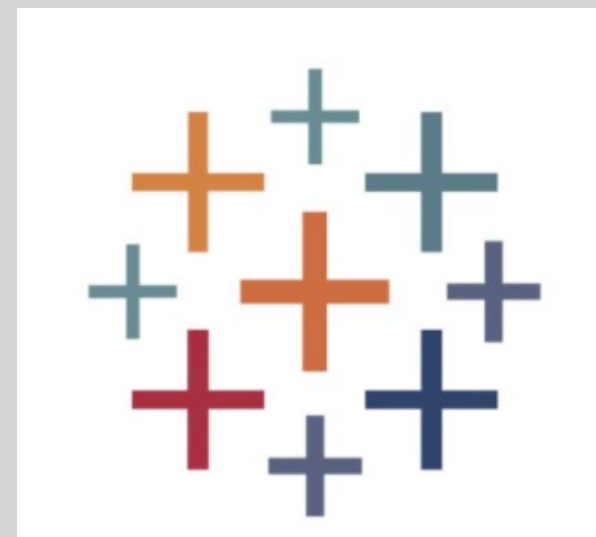


PowerBI



Data Portal

SaaS



Tableau

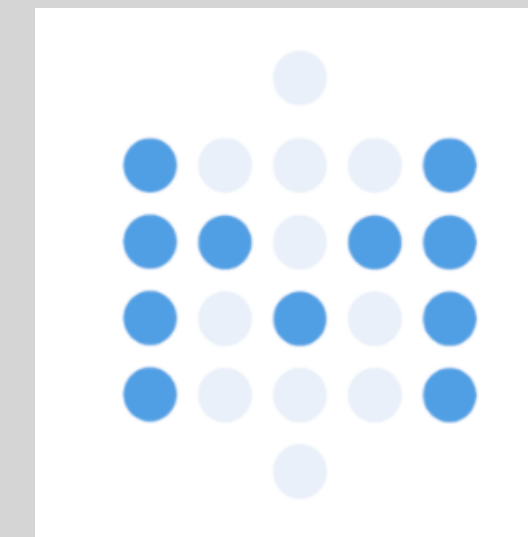


Looker

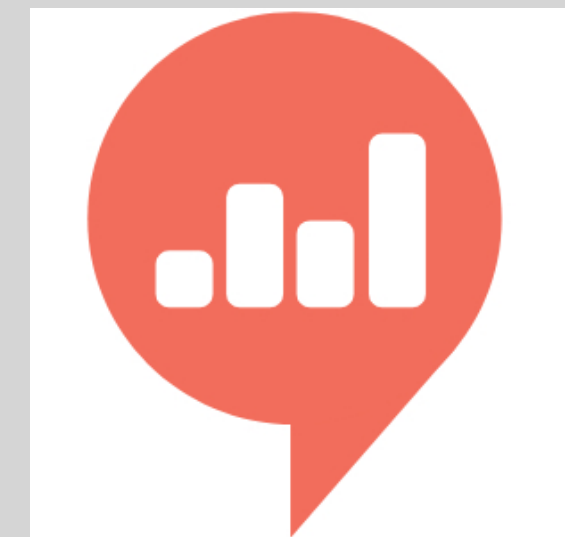


MotionBoard

OSS



Metabase

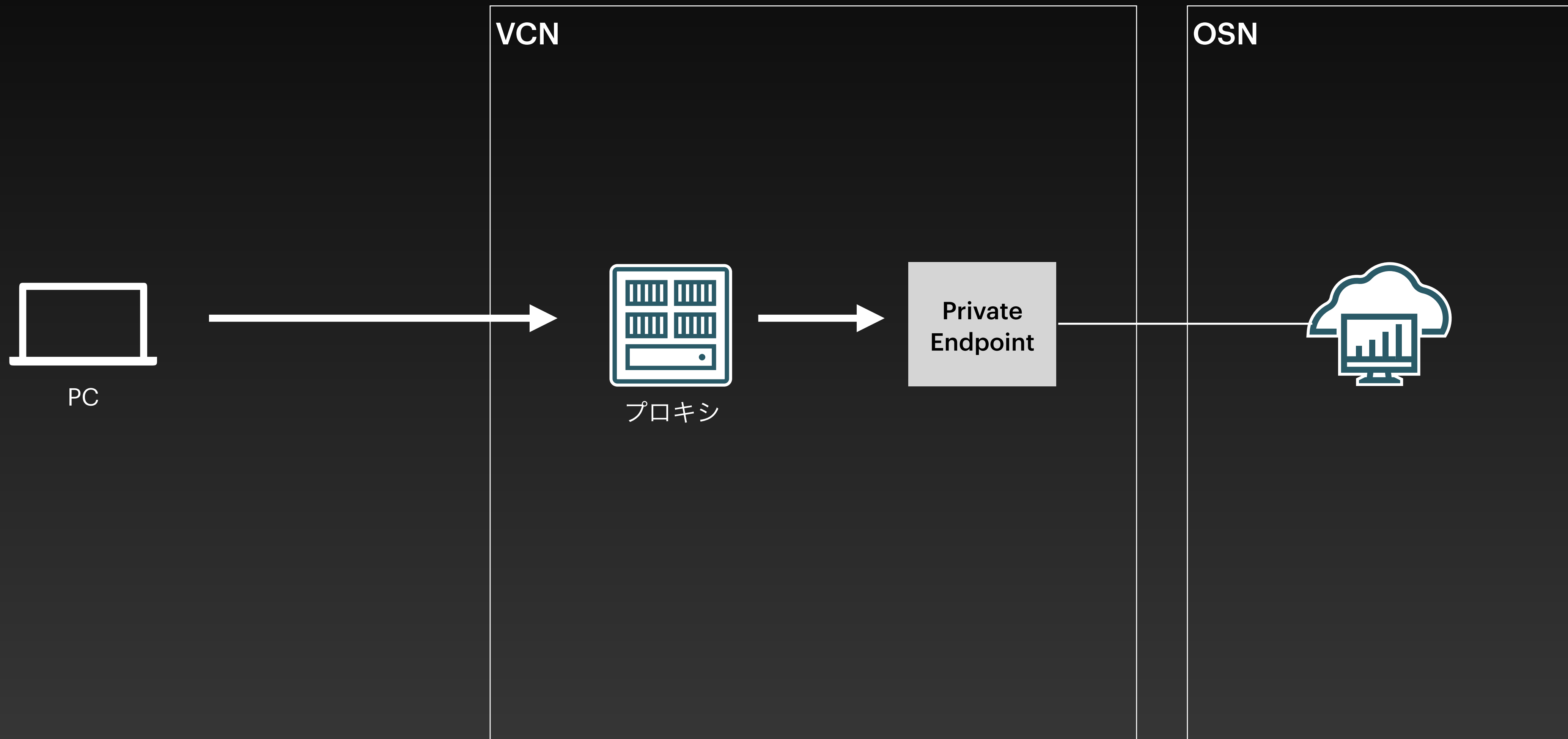


Redash

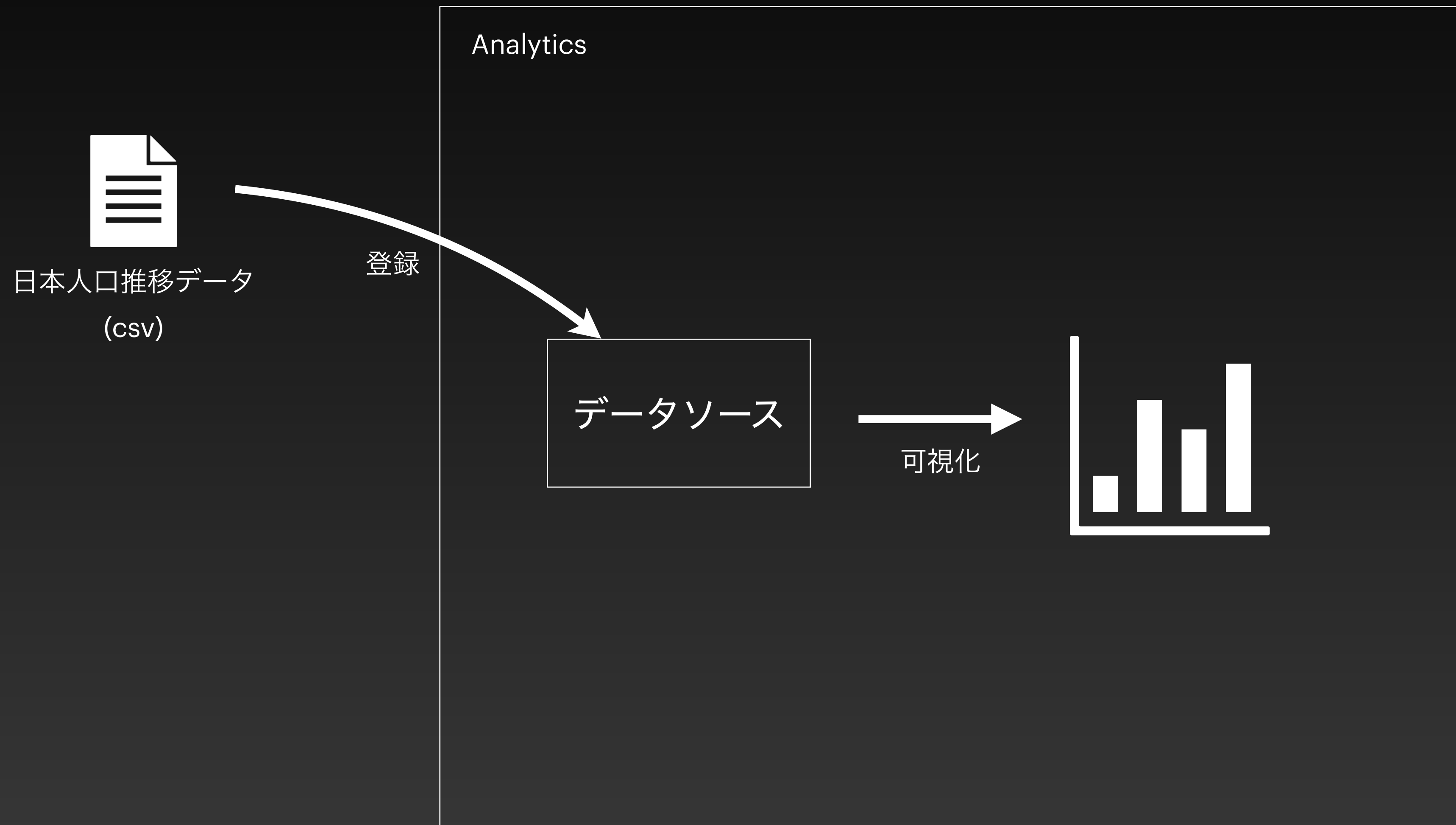
可視化・分析が簡単にできる



デモ構成

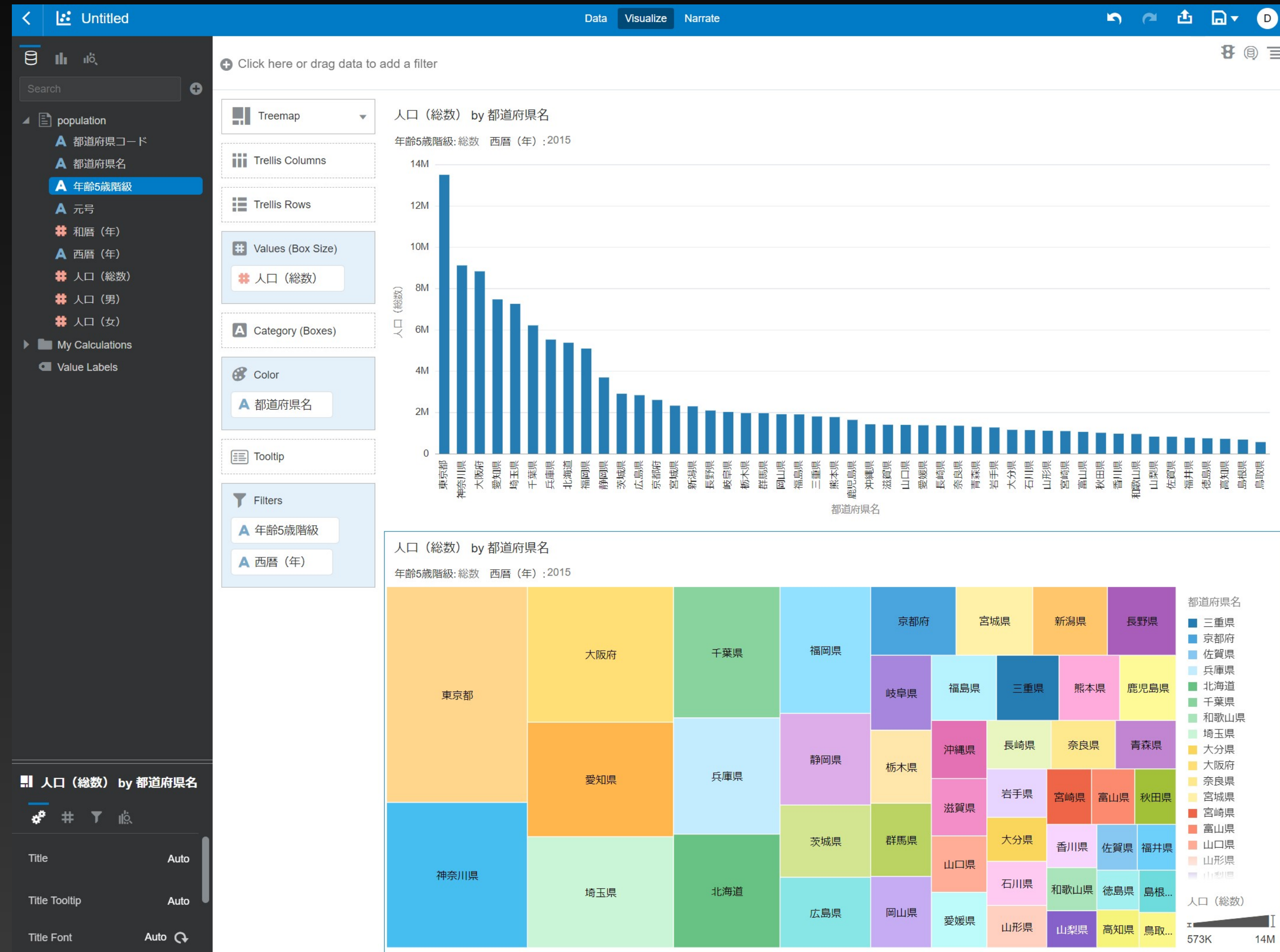


デモ① データ可視化



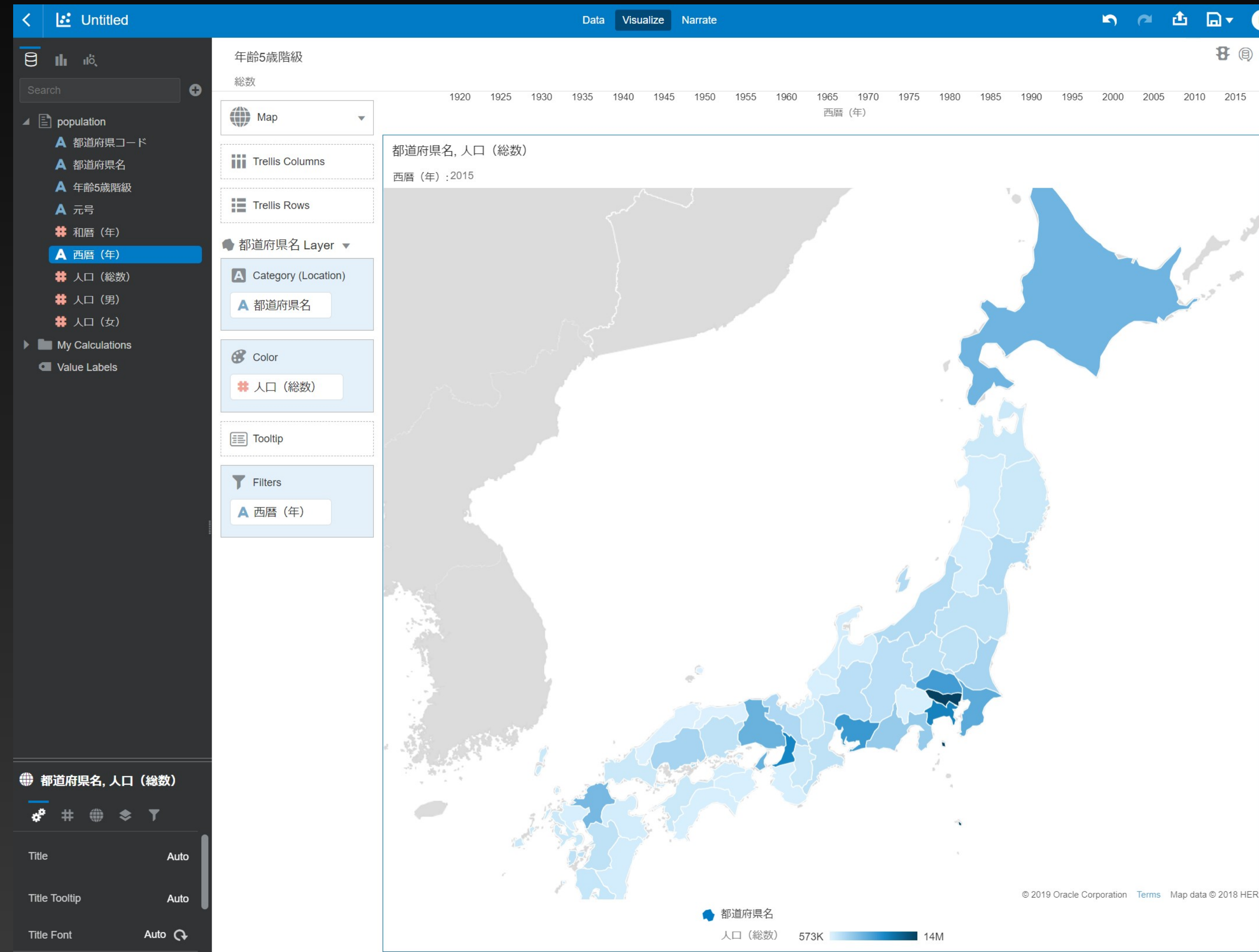
デモ① データ可視化

都道府県別人口データの グラフ化

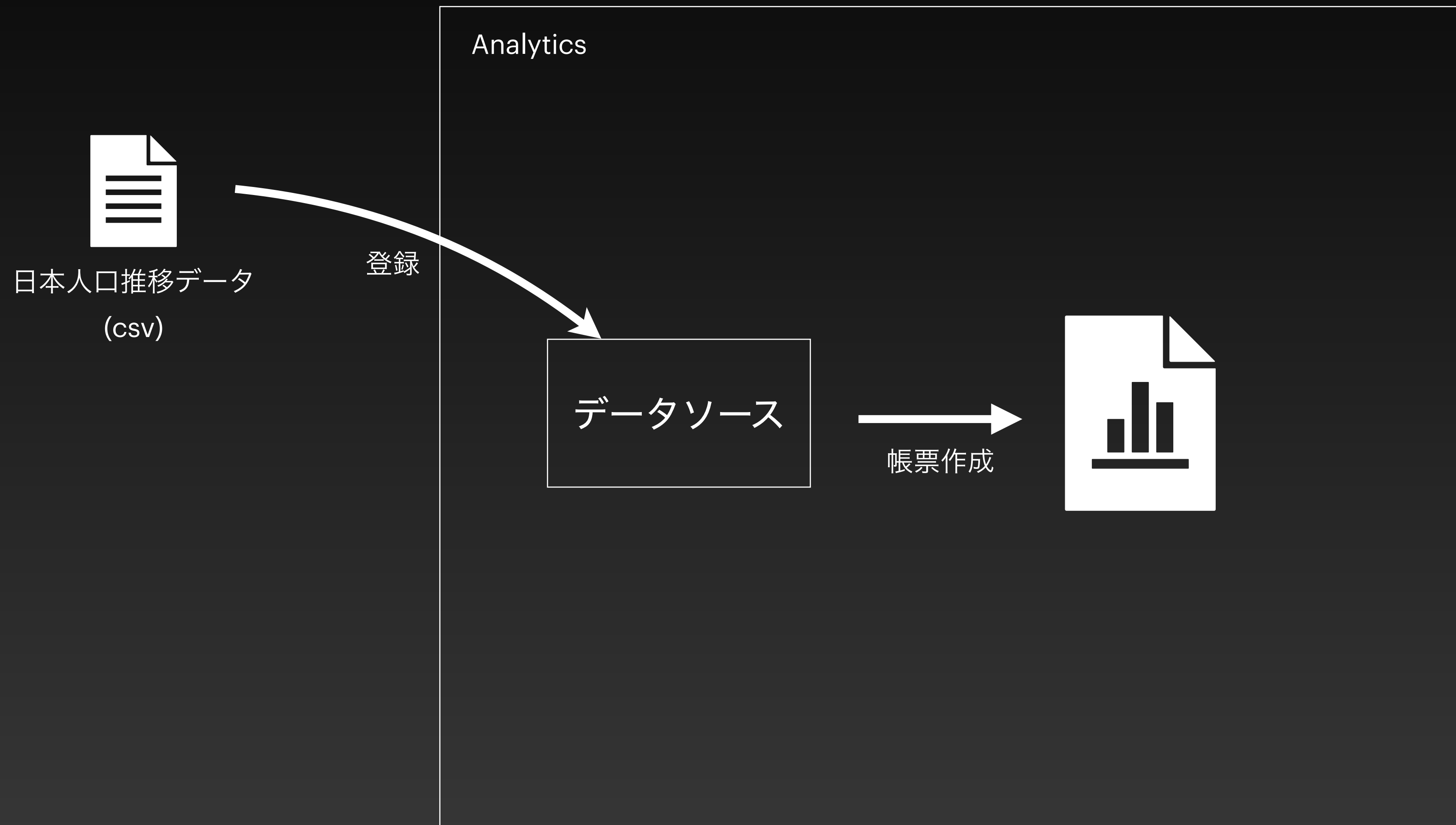


デモ① データ可視化

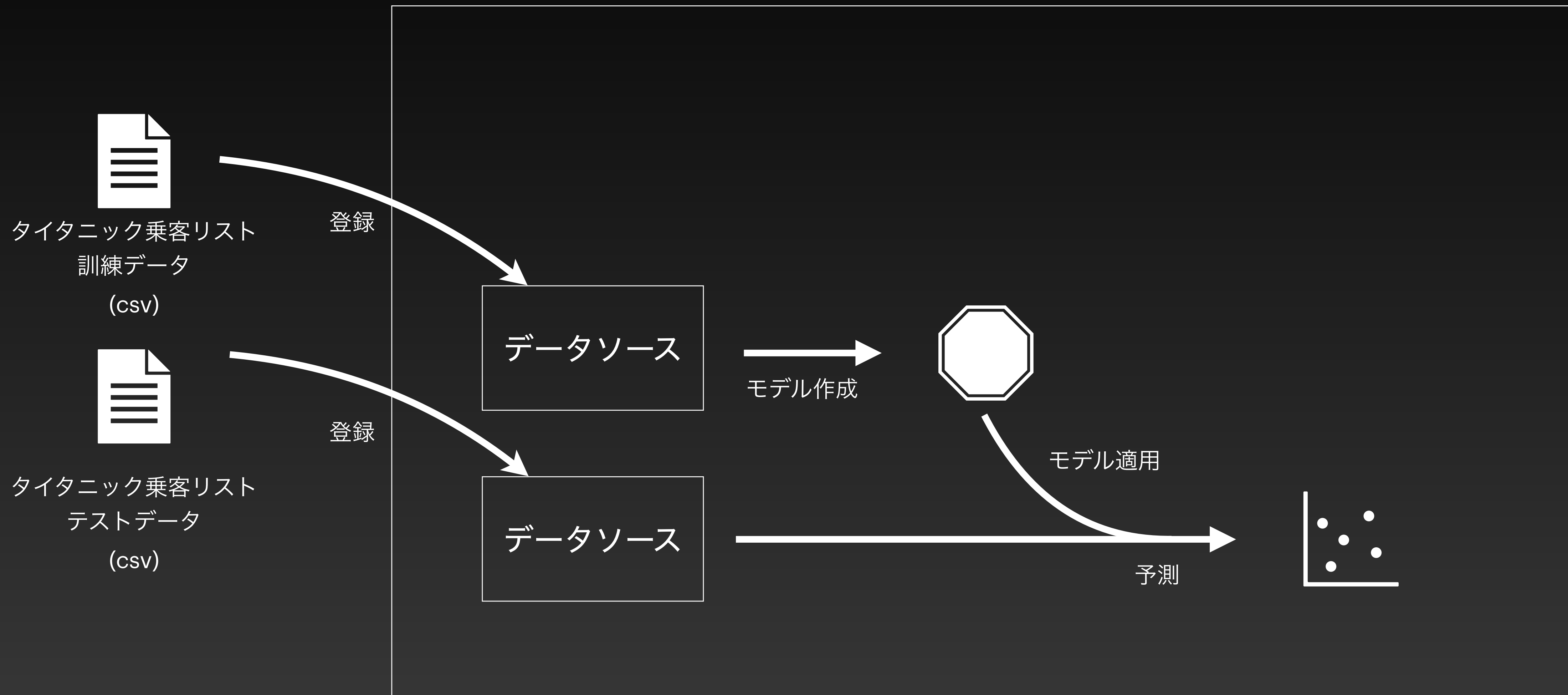
人口データを
地図データにマッピング



デモ② レポートテイニング



デモ③ 機械学習



デモ③ 機械学習

予測モデルの作成

The screenshot shows a software interface for training a machine learning model. The top navigation bar is blue and contains a back arrow, a double arrow, and the text "Train". On the right side of the bar are icons for undo, redo, play, save, and a user profile icon labeled "D". Below the bar is a search bar and a list of data manipulation actions: Add Data, Join, Union Rows, Filter, Aggregate, Save Data Set, Create Essbase Cube, Add Columns, Select Columns, Rename Columns, Transform Column, Merge Columns, Split Columns, Bin, Group, Branch, Cumulative Value, Time Series Forecast, and Analyze Sentiment. The main workspace is divided into two sections. The top section is a workflow diagram with three steps: "train", "Train Binary Classifier", and "Save Model". The bottom section is titled "Train Binary Classifier" and contains configuration options for a "Random Forest for model training" script. The target is set to "Survived". The positive class in the target is "Yes". The number of trees is 10, the sample size for a tree is 500, and the number of features for a tree is 3. Below the configuration is a data table with columns: Survived, Pclass, Sex, Age, SibSp, Parch, Fare, and Embarked. The table contains 15 rows of data.

Model Training Script: Random Forest for model training

* Target: Survived

target, the target(label) to learn/predict

Positive Class in Target: Yes

Positive class in the target value. Default is Yes.

Number Of Trees: 10

Number Of Trees in the forest

Sample Size For A Tree: 500

Sample size for a tree

Number Of Features For A Tree: 3

Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked
0	3	male	22	1	0	7.25	S
1	1	female	38	1	0	71.2833	C
1	3	female	26	0	0	7.925	S
1	1	female	35	1	0	53.1	S
0	3	male	35	0	0	8.05	S
0	3	male	30	0	0	8.4583	Q
0	1	male	54	0	0	51.8625	S
0	3	male	2	3	1	21.075	S
1	3	female	27	0	2	11.1333	S
1	2	female	14	1	0	30.0708	C
1	3	female	4	1	1	16.7	S
1	1	female	58	0	0	26.55	S
0	3	male	20	0	0	8.05	S
0	3	male	39	1	5	31.275	S

デモ③ 機械学習

モデルの適用

The screenshot displays a software interface for applying a machine learning model. The top navigation bar shows 'Untitled' and various utility icons. A search bar is present in the top left. A sidebar on the left contains a list of actions: Add Data, Join, Union Rows, Filter, Aggregate, Save Data Set, Create Essbase Cube, Add Columns, Select Columns, Rename Columns, Transform Column, Merge Columns, Split Columns, Bin, Group, Branch, Cumulative Value, Time Series Forecast, Analyze Sentiment, Train Numeric Prediction, Train Multi-Classifer, Train Clustering, Train Binary Classifier, and Apply Model. The main workspace shows a workflow with a 'test' node, an 'Apply Model' node, and a 'Save Data' button. Below this, the 'Save Data Set' configuration panel is visible, showing the following settings:

- Data Set: Untitled1
- Description: (empty)
- Save data to: Data Set Storage
- When Run: (empty)
- Prompt to specify Data Set:

The 'Columns' table in the configuration panel is as follows:

Name	Treat As	Default Aggregation
PassengerId	Measure	Sum
Pclass	Measure	Sum
Name	Attribute	
Sex	Attribute	
Age	Measure	Sum
SibSp	Measure	Sum

Below the configuration panel, a data table is displayed with the following columns: SibSp, Parch, Ticket, Fare, Cabin, Embarked, PredictedValue, PredictionConfidence, and PredictionGroup. The data rows are as follows:

SibSp	Parch	Ticket	Fare	Cabin	Embarked	PredictedValue	PredictionConfidence	PredictionGroup
0	0	330911	7.8292		Q	0.0	0.754169336160951	Sex IN ['male']
0	0	363272	7		S	0.09209952832979966	0.713613153180623	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8
0	0	240276	9.6875		Q	0.0	0.753601531297357	Sex IN ['male']
0	0	315154	8.6625		S	0.0	0.754169336160951	Sex IN ['male']
1	1	3101298	12.2875		S	0.38155927795097005	0.706834602586115	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8
0	0	7538	9.225		S	0.0	0.754169336160951	Sex IN ['male']
0	0	330972	7.6292		Q	0.199713073619249	0.740314797328201	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8
1	1	248738	29		S	0.4609051465995004	0.766405111653604	Sex IN ['male']
0	0	2657	7.2292		C	0.20466597238510237	0.717928384102721	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8
2	0	A/4 48871	24.15		S	0.0	0.754169336160951	Sex IN ['male']
0	0	349220	7.8958		S	0.0	0.754169336160951	Sex IN ['male']
0	0	694	26		S	0.2673703041073386	0.747197761751084	Sex IN ['male']
0	0	21228	82.2667	B45	S	0.9037444279171906	0.744835324128129	Sex NOT IN ['male']; Pclass <= 2.0
0	0	24065	26		S	0.07860162902011336	0.749916734121247	Sex IN ['male']
0	0	W.E.P. 5734	61.175	E31	S	0.9037444279171906	0.744835324128129	Sex NOT IN ['male']; Pclass <= 2.0
0	0	SC/PARIS 2167	27.7208		C	0.3854493756901684	0.736016766228875	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8
0	0	233734	12.35		Q	0.0	0.753601531297357	Sex IN ['male']
0	0	2692	7.225		C	0.0	0.736260205580568	Sex IN ['male']
0	0	STON/O2. 3101270	7.925		S	0.09209952832979966	0.713613153180623	Pclass > 1.0 AND Pclass <= 2.0; Age > 8.9 AND Age <= 16.8

さいごに

- ✓ OCIサービス中のAnalyticsの位置づけが何となく分かること
 - DWHに貯めたデータの分析や可視化ができる
(データ収集などは別サービス)
- ✓ Analyticsを使うと何ができるのか何となく分かること
 - 可視化・レポートニング・機械学習など
- ✓ 業務の中でAnalyticsの使い所が何となくイメージできること
 - システム稼働統計情報の分析
費用請求処理の自動化