Overview

- Our case study will focus on an anonymous regional chain of retail banks (XYZ Bank) operating primarily in the Tri-State Area and southern New England
- This mid-size financial services corporation boasts around 110 branches across 5 states and an annual revenue of approximately $750 million
- C-suite's currently stated goal for the first half of this decade is to expand operations to Mid-Atlantic states and Washington, DC as well as to begin marketing its new cryptocurrency loan offerings to Millennial and Gen-Z customers
- Nevertheless, Bank X still possesses an extremely outdated legacy BI system, with MS SQL server as the database backend but disjointed reporting and lack of data knowledge across different teams of the organization

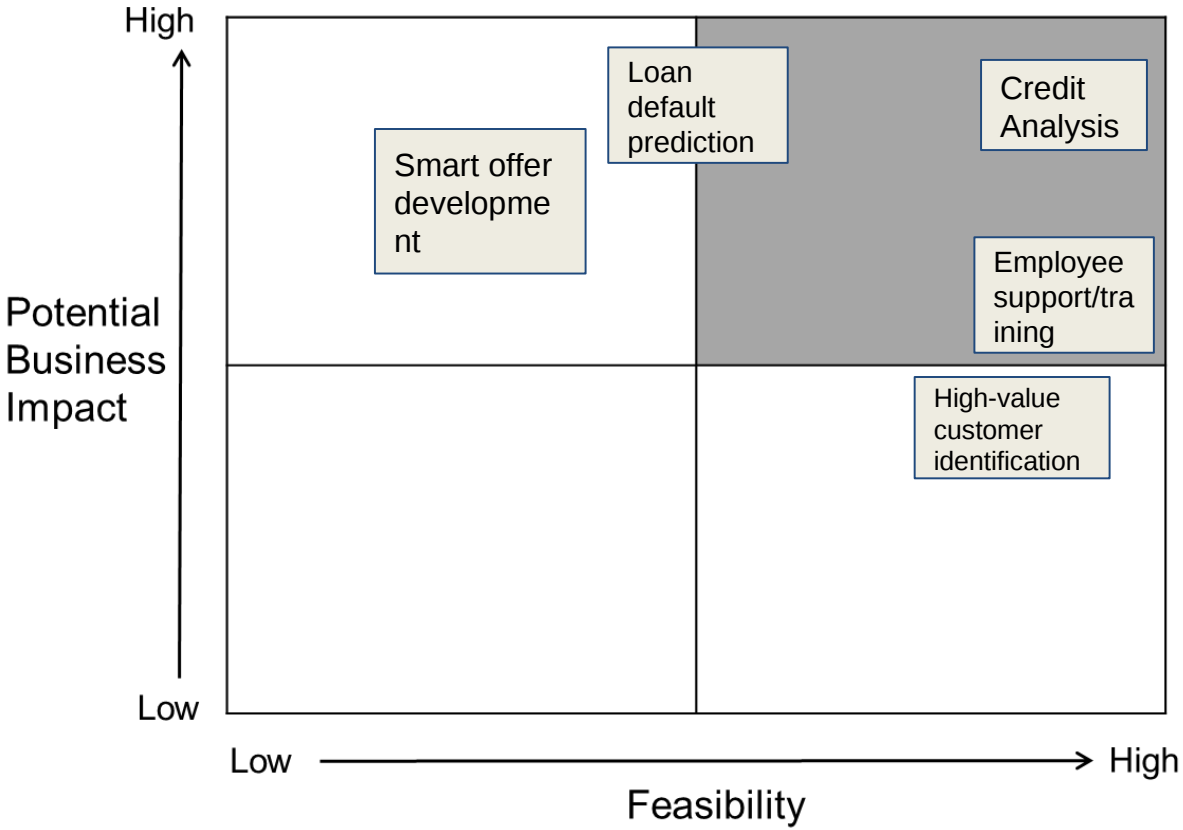
Opportunity Matrix

Business Process / Event	Organization/Workgroup: Customer Service	Human Resources	Data Science/IT	Risk Management	Marketing
Credit analysis				X	X
High-value customer identification	X				X
Loan default prediction			X	X	
Employee support/training		X	X		
Smart offer development				X	X

High-level Bus Matrix

Business Process / Event	Common Dimensions: Date	Customer	Loan	Bank Account	Credit Card	Branch	Employee	Offer Type
Credit analysis	X	X	X	X	X			X
High-value customer identification	X					X		
Loan default prediction	X							
Employee support/training	X					X	X	
Smart offer development	X	X	X		X			X

Prioritization Grid



Detailed Bus Matrix

Business Process / Event	Fact Tables	Granularity	Facts	Date	Customer	Credit Card	Offer Type	Bank Account
Loan default prediction	Loan payment transactions	1 row for every loan payment transaction	Loan payment amount	X	X			X
	Checking account transaction	1 row for every checking account transaction	Checking account transaction amount	X	X			X
Smart offer development	Credit card applications	1 row for every credit card application	Applicant chosen terms Applicant salary	X	X	X	X	X
	Loan applications	1 row for every loan application	Applicant chosen terms Applicant salary	X	X		X	X
	Offer types	1 row per offer APR and terms previously offered by the bank	Offer APR Offer term Offer premium Offer initial payment	X		X	X	

Conformed Dimension

(1) Delivery Steps of Conformed Dimension

1) Standardization

The purpose of standardization is to make the data coding methods and data formats of different data sources from different branches to the same, to lay a foundation for the next data matching.

2) Matching and duplication

Two aspects of data matching: one is to match the identification of different data sources and different attributes of the same thing to make the data more perfect; The other is to identify the same data from different data sources as duplicates, to lay a foundation for filtering in the next step

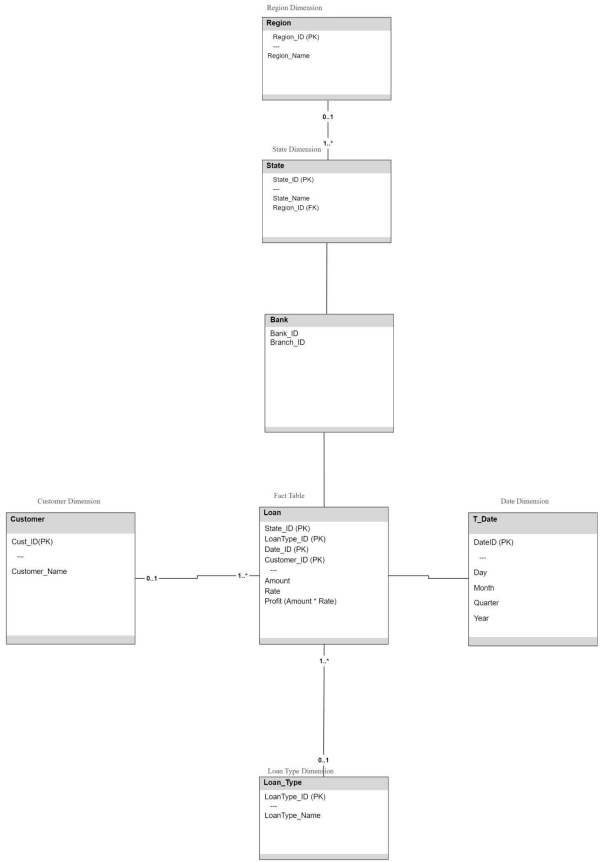
3) Surviving

Main purpose of data filtering is to select the conformed dimension as the master data, the final delivered conformed dimension data.

(2) Possible Conformed Dimension

- Date Dimension
- Customer Dimension
- Loan Dimension
- Credit Card Dimension
- Branch Dimension

Dimensional Model



Data Transformation

After extraction, the data needs to be transformed. Such as unifying data types, dealing with spelling errors, eliminating data ambiguity, and parsing into standard formats. Data transformation is usually the most complex part. It is also the longest step in ETL development. The range of data transformation is very wide. The process is to transform the data from simple data type to extremely complex data cleaning technology.

Data Transformation Rules:

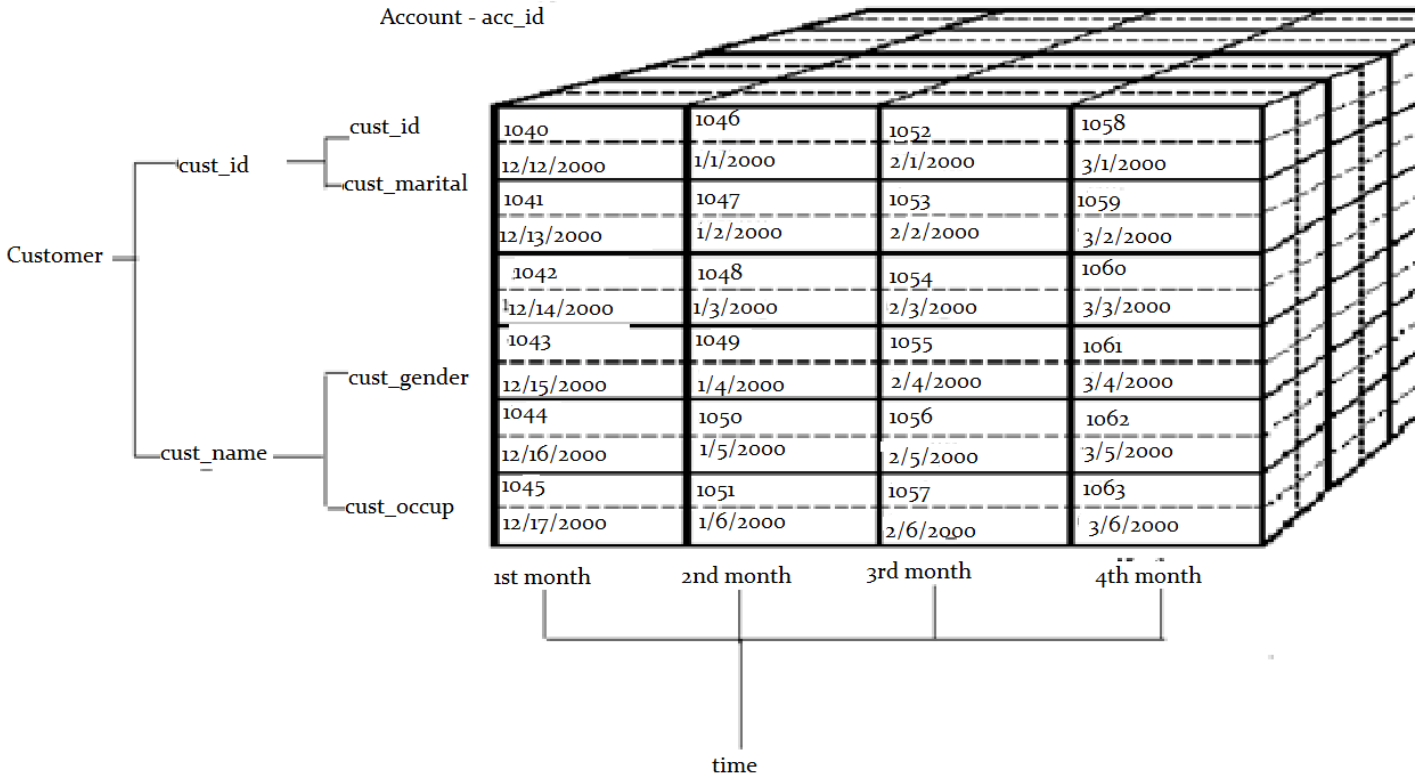
- Only specific data columns are loaded
- Unified coding
- Pre-calculation.
- Reordering improves query performance
- Merge data sources and remove duplicates
- Row column Transformation
- One column to multiple columns
- Merge duplicate columns
- Pre connection
- Data validation

Sample Two-Way Aggregate Table

Payments by Month by Customer Occupation Type

time_key	cust_occupation_key	count	total_revenue
546512	1006	6,363	$9,544,936.25
546512	1009	1,758	$11,504,352.78
546512	1011	353	$4,577,192.15
546118	1004	9,098	$10,924,696.40
546118	1003	23,544	$13,843,872.34
546087	1010	5,409	$7,677,899.22

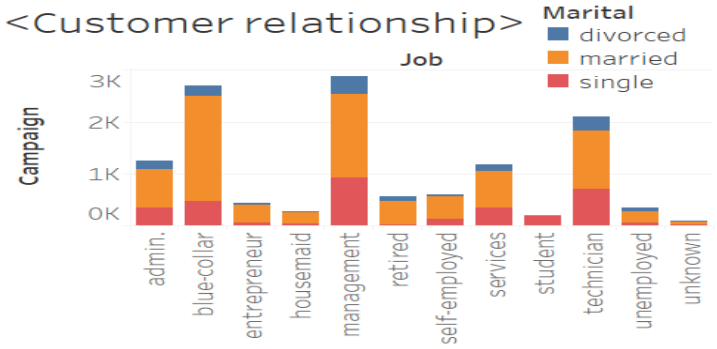
Data Cube

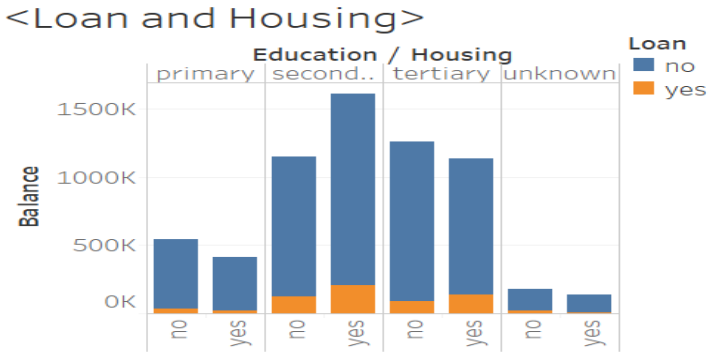

Account - acc_id
Customer
cust_id
cust_id
cust_marital
cust_name
cust_gender
cust_occup
1040
12/12/2000
1046
1/1/2000
1052
2/1/2000
1058
3/1/2000
1041
12/13/2000
1047
1/2/2000
1053
2/2/2000
1059
3/2/2000
1042
12/14/2000
1048
1/3/2000
1054
2/3/2000
1060
3/3/2000
1043
12/15/2000
1049
1/4/2000
1055
2/4/2000
1061
3/4/2000
1044
12/16/2000
1050
1/5/2000
1056
2/5/2000
1062
3/5/2000
1045
12/17/2000
1051
1/6/2000
1057
2/6/2000
1063
3/6/2000
1st month
2nd month
3rd month
4th month
time

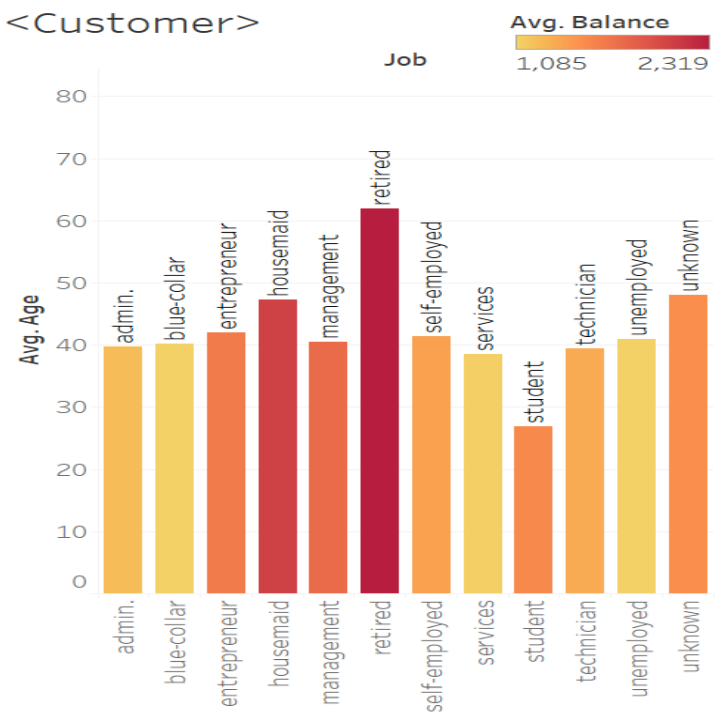
User/Task Analysis

User Role	Task	Delivery
Executives/Upper Management	Access high-level summaries to inform business strategy	Portal, reports
Knowledge workers: risk management/compliance staff, Data Science/quants	Ad-hoc queries for risk analysis, predictive modeling, revenue forecasting, etc	Portal, OLAP, MOLAP
Operational Workers: personal bankers, HR	Access to customer/employee data for everyday client-facing or administrative purposes	Portal, operational reports

Example visualization: Customer demographic dashboard in use by the marketing team

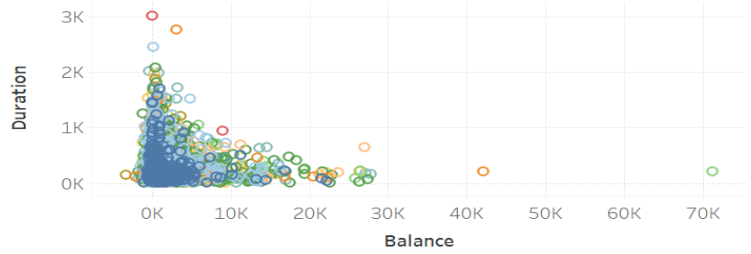




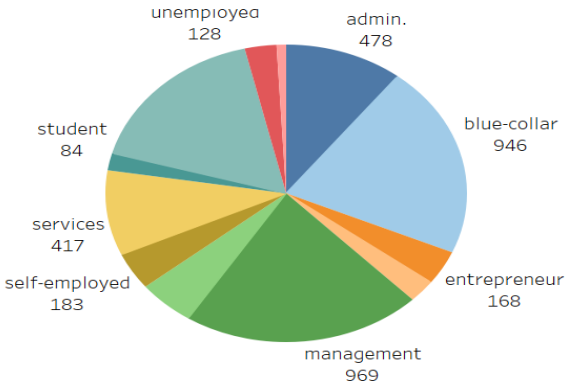


Example visualization: Customer demographic dashboard in use by the marketing team

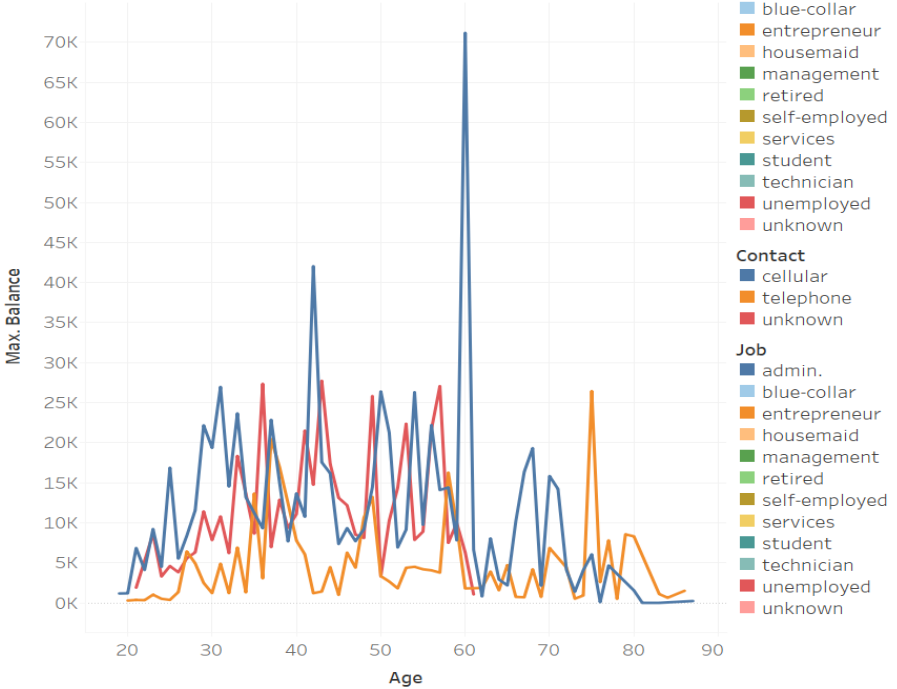
<Customer target>



<Customer Overview>

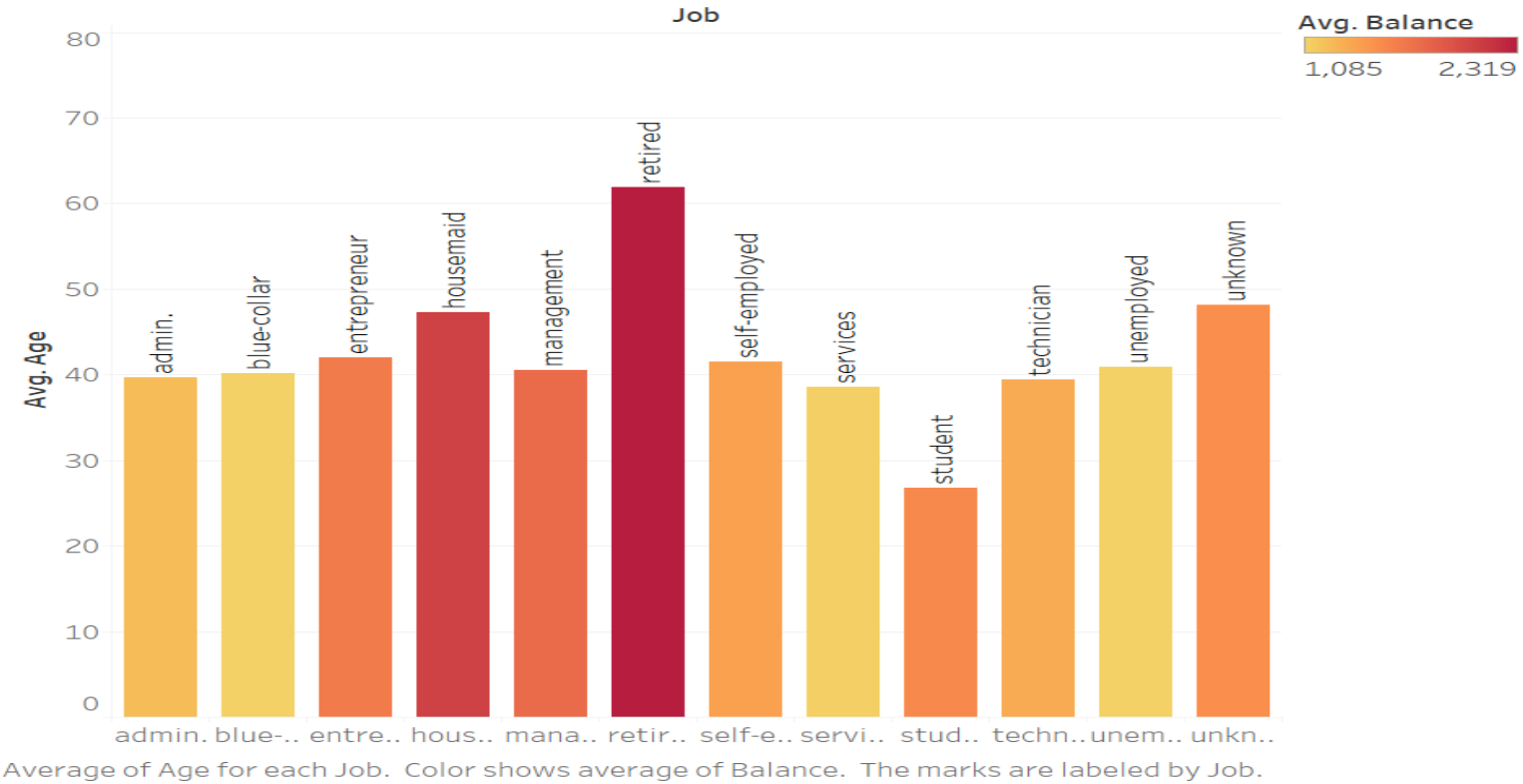


<Contact type>



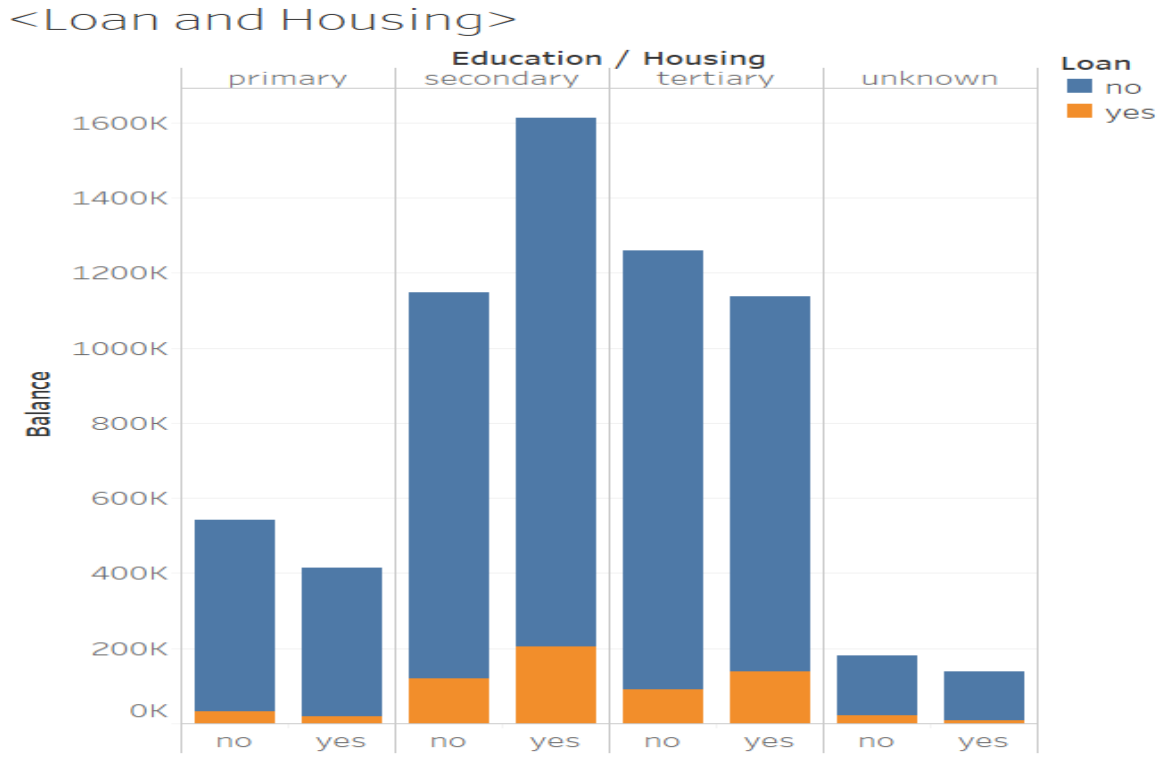
Customer Insights

<Customer>

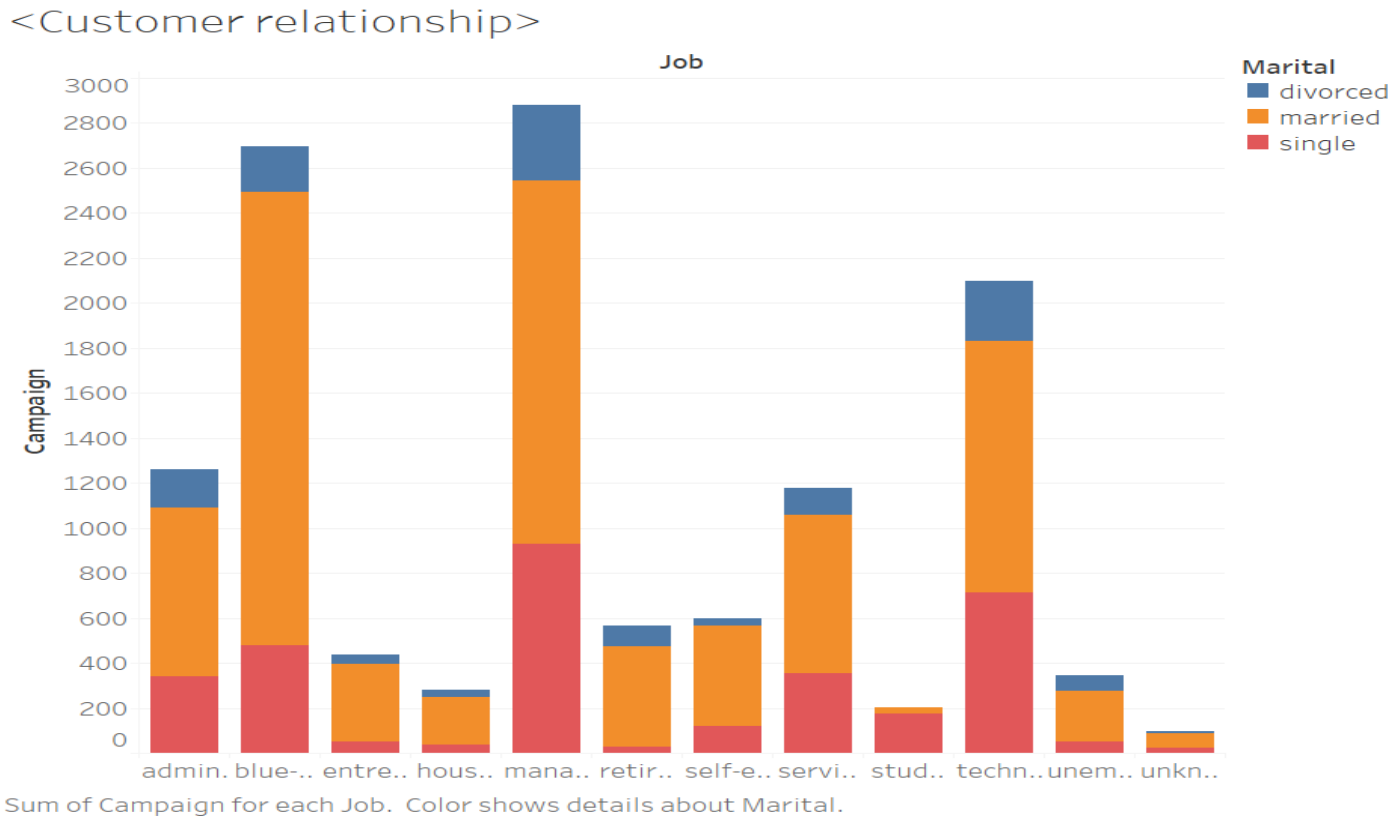


Average of Age for each Job. Color shows average of Balance. The marks are labeled by Job.

Education and Financial status

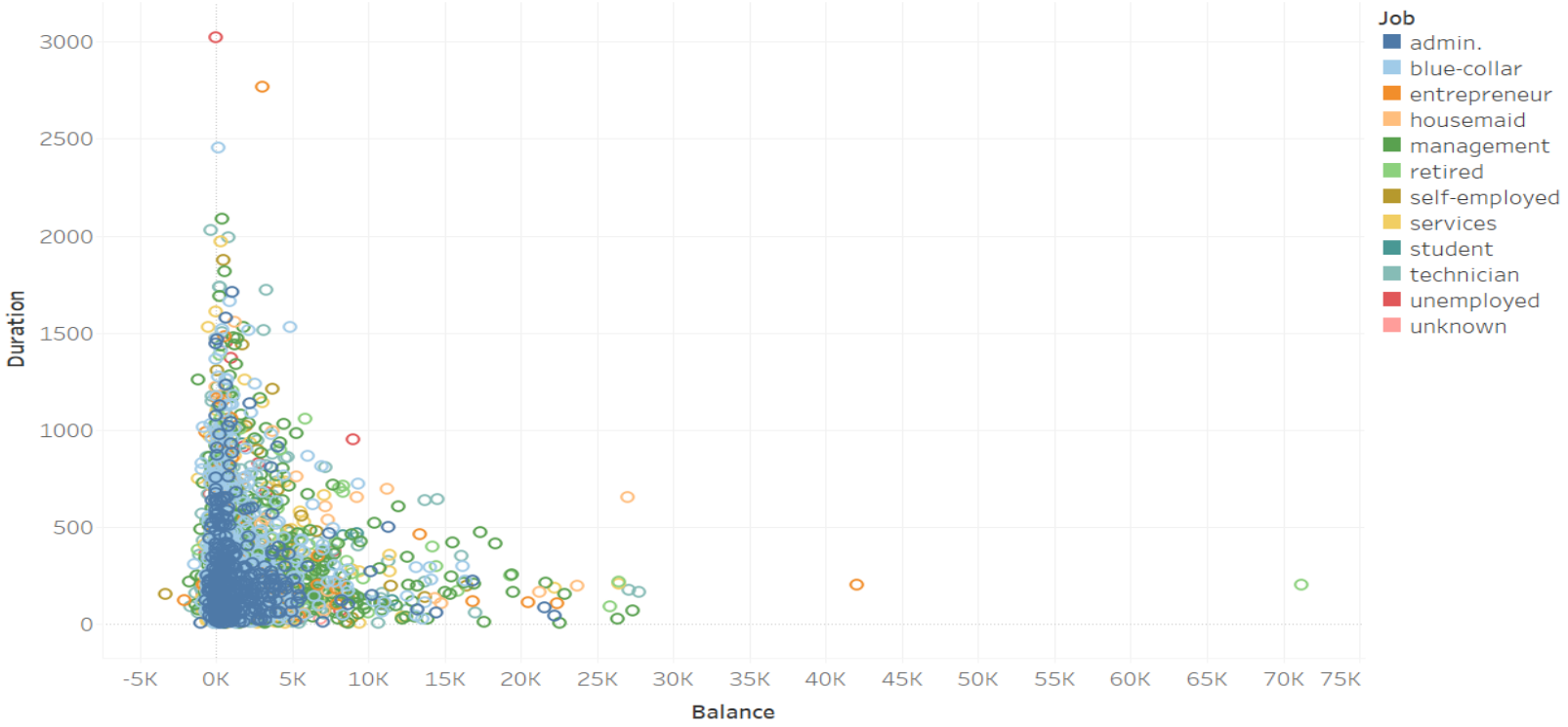


Campaign with job and marital status



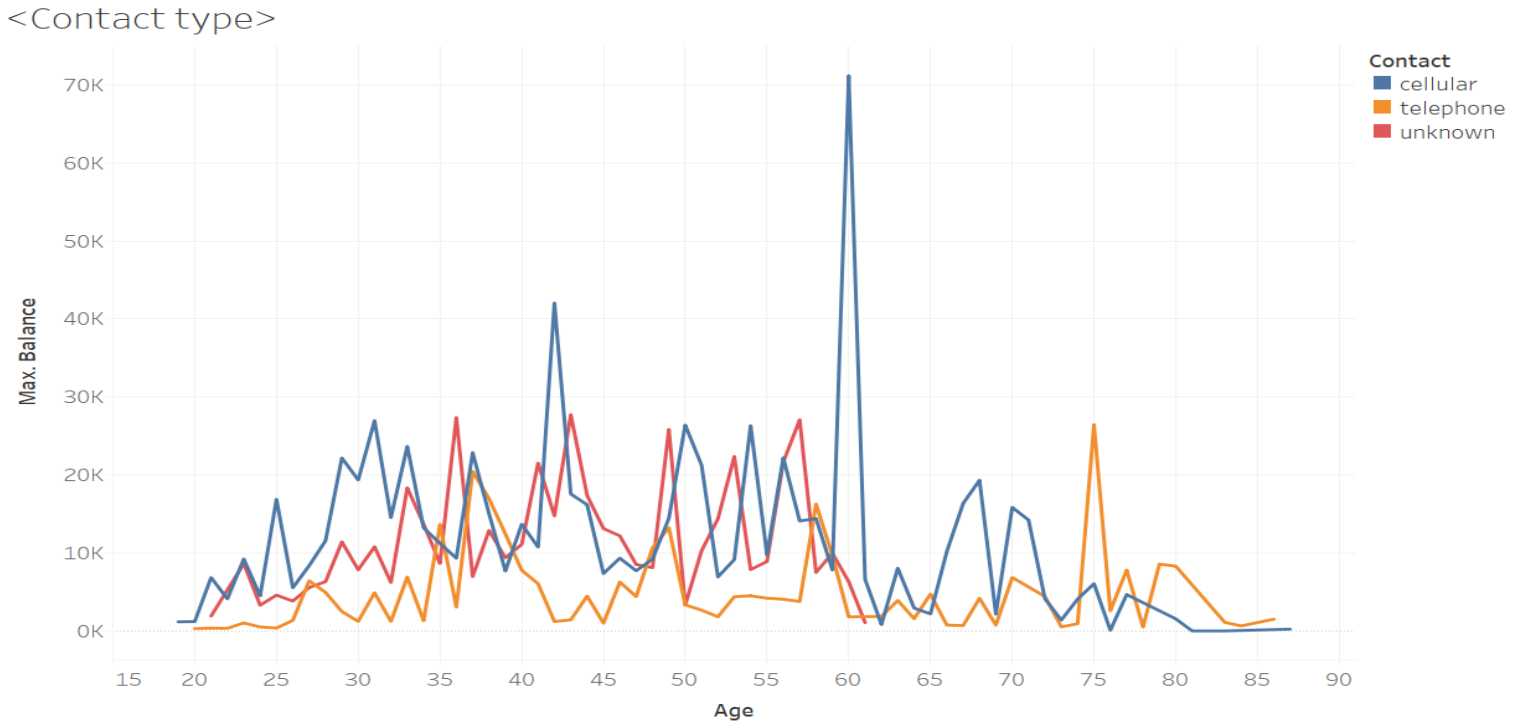
Target Insights

<Customer target>



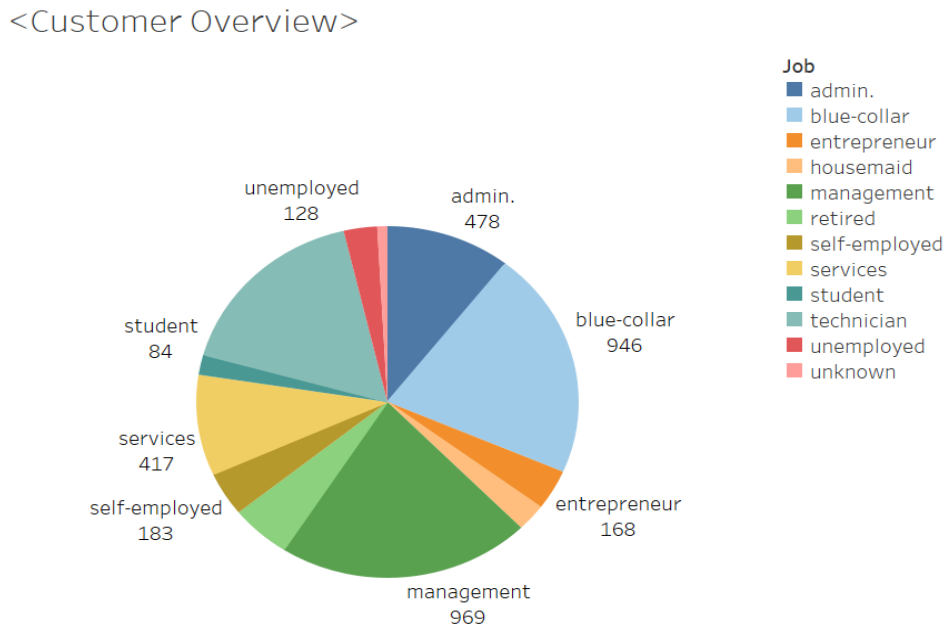
Balance vs. Duration. Color shows details about Job.

Contact type Overview



The trend of maximum of Balance for Age. Color shows details about Contact.

Customer Overview



Job and count of Job. Color shows details about Job. The marks are labeled by Job and count of Job.

Thank you!