



Data Sleek Data Warehouse - Clockify

Final Project Report

Group 1

Dexter Cai
Rose Feng
Jiaqi Li
Esther Chung
Yuhao Jiang
Lijie Li
Shuhan Lu

Table of Contents

| | |
|--|-----------|
| Executive Summary | 2 |
| Project Statement | 3 |
| Data Dictionary | 4 |
| ERD (Entity Relationship Diagram) | 6 |
| Data Literacy | 7 |
| Data Processing and Architecture | 9 |
| I. Time Tracking KPIs | 9 |
| II. Financial KPIs | 10 |
| III. Employee Performance KPIs | 11 |
| IV. Time Management KPIs | 11 |
| Visualization and Insights | 13 |
| I. Time Tracking KPIs | 13 |
| II. Financial KPIs | 14 |
| III. Employee Performance KPIs | 15 |
| IV. Time Management KPIs | 16 |
| Project Challenges | 18 |

Executive Summary:

Our overarching goal of this project is to help Data Sleek make the best use of its Clockify's data to empower more informed business decisions. We want to transform information sitting in the data warehouse into a solution that is flexible, intuitive, and actionable for the executives at Data Sleek. With this objective in mind, we have designed 8 KPIs (Key Performance Indicators) that we believe to be the most insightful ones, and have consolidated our visualizations and analyses into dynamic dashboards.

We want to help our clients carry out projects in a more efficient, cost-effective, productive, and organized manner, and that's why our KPIs target 4 main aspects: time-tracking, financials, employee performance, and time management, each with 2 meaningful metrics underneath.

One of our key findings is that data-science-related projects take significantly longer to complete than others. At a more granular level, implementation, optimization, and migration are the most timely tasks to complete on average, and these differences become even more drastic for certain clients. Interestingly, project management and data warehouse-related projects bring in the most money overall. On the other side, our time tracking dashboard and employee performance dashboard allows us to see and compare the productivity, performance, and contribution at the employee and department levels;

One can easily navigate through all of these aggregated metrics and granular metrics in our dashboards, as users can choose to sort, filter, and interact with them dynamically. It is definitely meaningful to find out the top-performing aspects of the projects, but we believe what's more valuable is perhaps to further analyze and diagnose the low-performing areas. This way we can help our clients easily locate their pain points and bottlenecks so that a more strategic decision can be made. Our dashboard can be both a solution and a starting point for stakeholders to analyze a problem of their interest. Regardless of how it will be used, what we can all agree on is that these data are all the incredibly valuable information that would otherwise be difficult to measure, and this is the key reason Data Sleek puts a tremendous emphasis on services like Clockify.

Project Statement:

In today's increasingly complex and hybrid business environment, companies are seeking new ways of tracking employees' time and efficiency with the use of data. With this goal in mind, many tools have emerged over the years, and Clockify is one of them. Clockify is a time-tracking software that lets companies monitor hours across projects, groups, and employees. Employees can start and stop the timer as they work, log their weekly hours in a timesheet, clock in from their device, and many more, all from Clockify's app or website. It can be integrated with other apps like Quickbooks and Jira, and work collectively to help companies improve their productivity and clarity of project management.

Data Slick is a data management and consulting company that makes extensive use of Clockify. As a company that aims to unleash the power of data and provide data solutions to clients worldwide, one of the key challenges is to figure out how to organize data, analyze the information, and extract valuable business insights from its data sources such as Clockify.

Our team, consisting of 7 data experts, has partnered with Data Slick to help drive the most value from Clockify's data. We have queried data from Data Slick's cloud data warehouse using Snowflake, analyzed and identified 8 KPIs (Key Performance Indicators), and put together a customizable dashboard displaying real-time information using Tableau, all of which will be explained in much more detail below. We believe that our solution will provide the convenience, flexibility, and clarity executives need to get a high-level understanding of Clockify's data, and consequently, help Data Slick and its clients make more strategic business decisions.

Data Dictionary (3 - Data Dict):

We created or made use of the following 18 tables in the database:

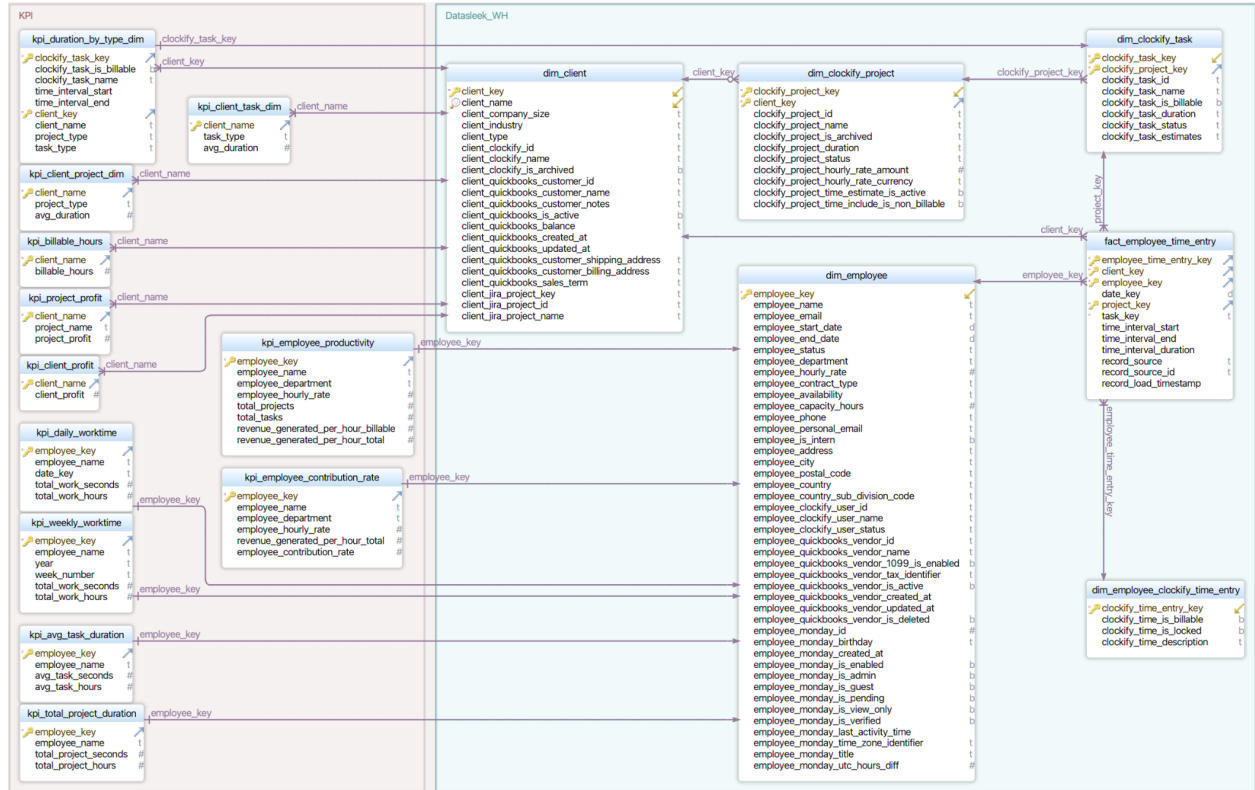
| Table | # Columns | # Records | Description |
|--------------------------------|-----------|-----------|--|
| DIM_CLIENT | 21 | 201 | Clients' high-level information, including their corresponding Clockify, Jira, and Quickbooks information |
| DIM_CLOCKIFY_PROJECT | 10 | 327 | Project-level information tracked by Clockify, including its time, hourly rate, client, and more |
| DIM_CLOCKIFY_TASK | 8 | 1,767 | Task-level information tracked by Clockify. Below Clockify's projects in terms of hierarchy |
| DIM_CLOCKIFY_TIME_ENTRY | 5 | 26,069 | Employees' time entry data for Clockify's projects and tasks |
| DIM_EMPLOYEE | 43 | 98 | Exhaustive table of employee's information, including demographics, credentials, and more |
| FACT_EMPLOYEE_TIME_ENTRY | 12 | 26,069 | Pulling the above information together and displaying the "time entry"-level information in much more details |
| KPI_DURATION_BY_TYPE_DIM | 9 | 17,985 | KPI to examine the average time of each project by its types (Project Management, Database Administration, Database Architecture, Data Warehouse, Data Lake, Data Visualization, Web, Engineering, Optimization) |
| KPI_CLIENT_PROJECT_DIM | 3 | 210 | KPI to examine the average project duration of each client by types (same as KPI_DURATION_BY_TYPE_DIM) |
| KPI_CLIENT_TASK_DIM | 3 | 425 | KPI to examine the average task duration of each client by types (same as KPI_DURATION_BY_TYPE_DIM) |
| KPI_BILLABLE_HOURS | 2 | 48 | Total billable hours of each client |
| KPI_PROJECT_PROFIT | 3 | 257 | Total Profit earned from each project |
| KPI_CLIENT_PROFIT | 2 | 52 | Total profit that each client earned from projects |
| KPI_AVG_TASK_DURATION | 4 | 85 | Average time each employee spent on tasks |
| KPI_TOTAL_PROJECT_DURATION | 4 | 85 | Total time each employee spent on tasks |
| KPI_DAILY_WORKTIME | 5 | 6,337 | Total time spent by each employee per day |
| KPI_WEEKLY_WORKTIME | 6 | 1,726 | Total time spent by each employee per week |
| KPI_EMPLOYEE_PRODUCTIVITY | 8 | 85 | Amount of work completed by employees in a given time period |
| KPI_EMPLOYEE_CONTRIBUTION_RATE | 6 | 85 | Hourly revenue the employee generated vs. the employee's hourly rate |

Here are the key columns that we made use of in designing our KPIs:

| Name | Table | Type | Description |
|---------------------------|----------------------|-------------------|--|
| CLIENT_KEY | DIM_CLIENT | VARCHAR(32) | Uniquely identifies a client |
| CLIENT_NAME | DIM_CLIENT | VARCHAR(16777216) | Name of a client |
| CLOCKIFY_PROJECT_KEY | DIM_CLOCKIFY_PROJECT | VARCHAR(32) | Uniquely identifies a project; matched with PROJECT_KEY |
| CLOCKIFY_PROJECT_NAME | DIM_CLOCKIFY_PROJECT | VARCHAR(16777216) | Name of a project (we used this to categorize project types) |
| CLOCKIFY_TASK_IS_BILLABLE | DIM_CLOCKIFY_TASK | BOOLEAN | Indicating whether the client will be charged for this task |
| CLOCKIFY_TASK_KEY | DIM_CLOCKIFY_TASK | VARCHAR(32) | Uniquely identifies a task; matched with TASK_KEY |
| CLOCKIFY_TASK_NAME | DIM_CLOCKIFY_TASK | VARCHAR(16777216) | Name of a task (we used this to categorize task types) |

| | | | |
|-------------------------------------|--------------------------------|-------------------|---|
| EMPLOYEE_DEPARTMENT | DIM_EMPLOYEE | VARCHAR(16777216) | Department of an employee |
| EMPLOYEE_HOURLY_RATE | DIM_EMPLOYEE | VARCHAR(16777216) | Hourly rate at which an employee is paid at |
| EMPLOYEE_KEY | FACT_EMPLOYEE_TIME_ENTRY | VARCHAR(32) | Uniquely identifies an employee |
| EMPLOYEE_NAME | DIM_EMPLOYEE | VARCHAR(16777216) | Name of an employee |
| PROJECT_KEY | FACT_EMPLOYEE_TIME_ENTRY | VARCHAR(16777216) | Uniquely identifies a project; matched with CLOCKIFY_PROJECT_KEY |
| TASK_KEY | FACT_EMPLOYEE_TIME_ENTRY | VARCHAR(16777216) | Uniquely identifies a task; matched with CLOCKIFY_TASK_KEY |
| TIME_INTERVAL_DURATION | FACT_EMPLOYEE_TIME_ENTRY | NUMBER(18,0) | Duration of a time interval; duration between start time and end time |
| TIME_INTERVAL_END | FACT_EMPLOYEE_TIME_ENTRY | TIMESTAMP_NTZ(9) | Timestamp of when an employee "punches out" |
| TIME_INTERVAL_START | FACT_EMPLOYEE_TIME_ENTRY | TIMESTAMP_NTZ(9) | Timestamp of when an employee "punches in" |
| PROJECT_TYPE | KPI_DURATION_BY_TYPE_DIM | VARCHAR(16777216) | Categorizes the projects by specific words from project name |
| TASK_TYPE | KPI_DURATION_BY_TYPE_DIM | VARCHAR(16777216) | Categorizes the tasks by specific words from task name |
| AVG_DURATION | KPI_CLIENT_PROJECT_DIM | NUMBER(38,0) | Average duration for a specific task type |
| BILLABLE_HOURS | KPI_BILLABLE_HOURS | NUMBER(18,0) | Total billable hours of a client |
| PROJECT_PROFIT | KPI_PROJECT_PROFIT | VARCHAR(16777216) | Profit of a project |
| CLIENT_PROFIT | KPI_CLIENT_PROFIT | VARCHAR(16777216) | Profit of a client |
| AVG_TASK_SECONDS | KPI_AVG_TASK_DURATION | NUMBER(38,2) | Average time taken by an employee to complete a task in seconds |
| AVG_TASK_HOURS | KPI_AVG_TASK_DURATION | NUMBER(38,2) | Average time taken by an employee to complete a task in hours |
| TOTAL_PROJECT_SECONDS | KPI_TOTAL_PROJECT_DURATION | NUMBER(38,0) | Total time taken by an employee to complete a project in seconds |
| TOTAL_PROJECT_HOURS | KPI_TOTAL_PROJECT_DURATION | NUMBER(38,0) | Total time taken by an employee to complete a project in hours |
| TOTAL_WORK_SECONDS (D) | KPI_DAILY_WORKTIME | NUMBER(38,0) | Total work time of an employee in seconds per day |
| TOTAL_WORK_HOURS (D) | KPI_DAILY_WORKTIME | NUMBER(38,0) | Total work time of an employee in hours per day |
| YEAR | KPI_WEEKLY_WORKTIME | NUMBER(38,0) | Year scope for an employee's weekly performance |
| WEEK_NUMBER | KPI_WEEKLY_WORKTIME | NUMBER(38,0) | Week scope for an employee's weekly performance |
| TOTAL_WORK_SECONDS (W) | KPI_WEEKLY_WORKTIME | NUMBER(38,0) | Total work time of an employee in seconds per week |
| TOTAL_WORK_HOURS (W) | KPI_WEEKLY_WORKTIME | NUMBER(38,0) | Total work time of an employee in hours per week |
| TOTAL_PROJECTS | KPI_EMPLOYEE_PRODUCTIVITY | NUMBER(38,0) | Number of project done by an employee |
| TOTAL_TASKS | KPI_EMPLOYEE_PRODUCTIVITY | NUMBER(38,0) | Number of task done by an employee |
| REVENUE_GENERATED_PER_HOUR_BILLABLE | KPI_EMPLOYEE_PRODUCTIVITY | NUMBER(38,2) | Average revenue generated by an employee per billable hour |
| REVENUE_GENERATED_PER_HOUR_TOTAL | KPI_EMPLOYEE_PRODUCTIVITY | NUMBER(38,2) | Average revenue generated by an employee per hour (includes non-billable hours) |
| EMPLOYEE_CONTRIBUTION_RATE | KPI_EMPLOYEE_CONTRIBUTION_RATE | NUMBER(38,2) | Hourly revenue an employee generated vs. employee's hourly rate |

ERD (Entity Relationship Diagram):



Data Literacy:

In this section, we will describe the source of our data by its structure, size, and analytical inference. Clockify is a time-tracking software that helps individuals and teams improve productivity, efficiency, and billable hours by monitoring and recording their time spent on various tasks and projects.

The primary sources of data sources for this project are as follows:

- DIM_CLIENT
- DIM_EMPLOYEE
- DIM_CLOCKIFY_PROJECT
- DIM_CLOCKIFY_TASK
- DIM_EMPLOYEE_CLOCKIFY_TIME_ENTRY
- FACT_EMPLOYEE_TIME_ENTRY

These data sources include information about the time employees spend, costs that customers and Datasleek spend, and revenue that customers and Datasleek earned on various tasks and projects.

The amount of data in these sources may vary depending on the size of the client and the number of employees using Clockify software. As of the snapshot of the schema information from the dataset Datasleek shared with us, the size of DS_DWH_UCLA is approximately 5.4 Mb which contains 21 tables.

| Schema Name | Byte Size |
|-------------|-----------|
| DBT_DWH | 5,663,232 |
| DS_DWH_UCLA | 5,394,944 |

And for this project, we will focus on the following tables,

| Table Name | # Rows | Byte Size |
|--------------------------|--------|-----------|
| FACT_EMPLOYEE_TIME_ENTRY | 26,069 | 2,511,360 |
| DIM_CLOCKIFY_TIME_ENTRY | 26,069 | 1,347,072 |

| | | |
|----------------------|-------|---------|
| DIM_CLOCKIFY_TASK | 1,767 | 121,344 |
| DIM_EMPLOYEE | 98 | 33,792 |
| DIM_CLIENT | 201 | 32,768 |
| DIM_CLOCKIFY_PROJECT | 327 | 29,696 |

To extract meaningful KPIs out of these tables, the first step is to identify relevant data fields and variables that can be used to create meaningful KPIs that align with business goals, monitoring productivity and enhancing efficiency for employees. And the next step is to design distinctive dimensions for KPIs. Once the dimensions are established and relevant data fields are identified, we can proceed to restructure the dataset, building new tables for storing indicators that capture patterns, trends, and insights into business operations.

Overall, effective data analysis requires an in-depth understanding of data sources, business context, and analytical tools and techniques. By leveraging these resources, it is possible to gain valuable insights and make informed business decisions based on a designed KPI dashboard.

Data Processing and Architecture (KPI Queries):

In this section, we will discuss the process of scripting the SQL code for data preprocessing using Clockify data. The objective of this process is to extract relevant data and transform it into a format that can be used to calculate Key Performance Indicators (KPIs). The SQL code has been written to extract the columns required for calculating KPIs and categorize the tasks and projects based on specific keywords.

Categorizing these 8 different KPIs into 4 groups serves the purpose of organizing and better understanding key performance indicators related to time tracking, financials, employee performance, and time management. By grouping these KPIs together, it allows businesses to analyze and track different aspects of their operations in a more structured and comprehensive way. For example, time-tracking KPIs can help businesses monitor the time spent on different tasks, while financial KPIs can provide insights into profitability and billing practices. Meanwhile, employee performance KPIs can help track productivity and task completion rates, and time management KPIs can help identify areas where time is being spent inefficiently. By categorizing these KPIs, businesses can gain a clearer understanding of their performance across different areas, and make more informed decisions based on the insights gained.

1. Time Tracking KPIs [KPI1-1.sql](#) [KPI1-2.sql](#)

The first Time Tracking KPI is the amount of time employees spend at their workplace on a daily and weekly basis. Generally, it is calculated by first joining the `fact_employee_time_entry` table and `dim_employee_table` by `employee_key`, and then taking the sum of `time_interval_duration` field and grouping the sum of time duration by each `date_key` in the `fact_employee_time_entry` table and by each `employee_key` in the `dim_employee` table in our data source. To make our data source more structured and help companies make better decisions on each employee's performance, we created two result tables for the daily work time and weekly work time. One result table would display the employee key, employee name, date, and daily work time in seconds and hours; and the other result table would show the employee key, employee name, year, week of the year, and weekly work time in seconds and hours. This KPI is particularly important for our project because it shows some fundamental problems with employees' work efficiency and work habits, allowing us to dig deeper into these issues in the next steps. However, while working on the first Time Tracking KPI, we found that one employee, Franck Leveneur, may have entered the wrong working hours because his work start time was entered twice in the data



source for two days on 2021-07-12 and 2021-07-19, and his total daily work time exceeds 24 hours on these days. This is something we're looking to address with the company in the future.

The second Time Tracking KPI is the average time employees spend to complete a task and the total time they spend to complete a project. This KPI is also calculated by first joining the tables, which are `dim_clockify_task`, `dim_clockify_project`, `fact_employee_time_entry`, and `dim_employee`, based on `employee_key`, `clockify_project_key`, and `clockify_task_key`. Afterward, the average time per task KPI is calculated by taking the average of the `time_interval_duration` field for all the tasks and grouping the average of task time duration by each `employee_key` in the `dim_employee` table. The total time employees spend on a project is calculated by taking the sum of the `time_interval_duration` field for all the projects and grouping the sum of project time duration by each `employee_key` in the `dim_employee` table in our data source. This KPI allows us to understand the type of tasks and projects that employees usually spend the most time on on an employee level and help the companies to think of some personalized strategies they can enforce to help increase the work efficiency of their employees. Querying data for this KPI also builds a foundation for our analysis of the Time Management KPIs later.

2. Financial KPIs [KPI2-1.sql](#) [KPI2-2.sql](#)

The first part of Financial KPIs is about the project and client profitability. For a project's profit, we need to calculate its cost and revenue. A project's cost refers to the salaries of employees who worked on the project. Therefore, we calculate the hours each employee spent on each project using the `time_interval_duration` column from the `fact_employee_time_entry` table and multiply them with `employee_hourly_rate`, and then we sum the multiplications together grouped by project, thereby getting the costs of projects. In order to calculate each project's revenue, we calculate the hours of each project using the `time_interval_duration` and multiply them with `clockify_project_hourly_rate_amount`. After we obtain both the costs and revenues of projects, project profits are calculated as revenues minus costs. This KPI can help the company identify which projects are worth pursuing and which ones may need to be reevaluated.

For a client's profit, similarly, we need to calculate its cost and revenue. Its cost is also from the salaries of employees, so we use the same logic and method to calculate the costs. Clients' revenues are calculated from the `client_payment_amount` column from the `fact_client_payment` table. Therefore, clients' profits are calculated as revenues minus costs. This KPI can help us identify which clients are well profitable and which are not.

The second part of Financial KPIs is about client billable hours. Targeting billable tasks that are found by the where clause setting `clockify_task_is_billable = TRUE`, we calculate billable hours using `client_payment_amount` divided by `clockify_project_hourly_rate_amount` and sum it together based on different clients. This KPI measures the amount of time that is billed to the client on Clockify. It can help the company ensure that employees are spending their time on billable tasks and not on non-billable administrative tasks.

3. Employee Performance KPI [KPI3-1.sql](#) [KPI3-2.sql](#)

The first KPI we chose for measuring and evaluating employee performance is **employee productivity**, measured by how much revenue each employee made on average across all the projects they took part in. `REVENUE_GENERATED_PER_HOUR_BILLABLE` uses the total billable hours of each employee to calculate the average hourly revenue generated, while `REVENUE_GENERATED_PER_HOUR_TOTAL` uses the total hours (both billable and non-billable) as the denominator for the calculation. By looking at this KPI, Clockify can get a sense of how each employee is performing in the sense of generating revenue.

The second KPI is **employee contribution rate**, which measures the size of the employee's hourly generated revenue relative to their hourly wage. This KPI allows the company to understand the "cost-effectiveness" of hiring different employees and make future decisions on their human resource decisions. Given departmental information, we can also calculate the average contribution rate and average productivity of each department to get a general understanding of each department's contribution to revenue generation, as well as each employee's performance relative to the department they are in.

4. Time Management KPIs [KPI4-1.sql](#) [KPI4-2.sql](#)

The first step is to extract the relevant columns from the Clockify data that are required for calculating KPIs. The SQL code starts with a SELECT statement, which selects the `CLOCKIFY_TASK_KEY`, `CLOCKIFY_TASK_IS_BILLABLE`, `CLOCKIFY_TASK_NAME`, `TIME_INTERVAL_START`, `TIME_INTERVAL_END`, `TIME_INTERVAL_DURATION` as `DURATION`, `c.CLIENT_KEY`, and `c.CLIENT_NAME` columns.

Next, the projects are categorized into different types based on specific keywords using a CASE statement. The SQL code checks the `CLOCKIFY_PROJECT_NAME` column for specific keywords and categorizes the projects accordingly. For instance, if the project name contains the keyword 'Project

Management,' it is categorized as a 'Project Management' project. Similarly, if the project name contains the keyword 'Data Warehouse,' it is categorized as a 'Data Warehouse' project.

```
//Categorize the projects by specific words
CASE
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Project Management%' THEN 'Project Management'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Database Administration%' THEN 'Database Administration'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Database Architecture%' THEN 'Database Architecture'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Data Warehouse%' THEN 'Data Warehouse'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Data Lake%' THEN 'Data Lake'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Data Visualization%' THEN 'Data Visualization'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Web%' THEN 'Web'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Engineering%' THEN 'Engineering'
  WHEN CLOCKIFY_PROJECT_NAME LIKE '%Optimiz%' THEN 'Optimization'
  ELSE CLOCKIFY_PROJECT_NAME
END AS PROJECT_TYPE,
```

The same process is repeated for categorizing tasks based on specific keywords. The SQL code checks the CLOCKIFY_TASK_NAME column for specific keywords and categorizes the tasks accordingly. For instance, if the task name contains the keyword 'Communication,' it is categorized as a 'Communication' task. Similarly, if the task name contains the keyword 'Meeting,' it is categorized as a 'Meeting' task.

The SQL code then joins the tables to extract the necessary data. It joins the DIM_CLOCKIFY_TASK, FACT_EMPLOYEE_TIME_ENTRY, DIM_CLOCKIFY_PROJECT, and DIM_CLIENT tables using the relevant keys.

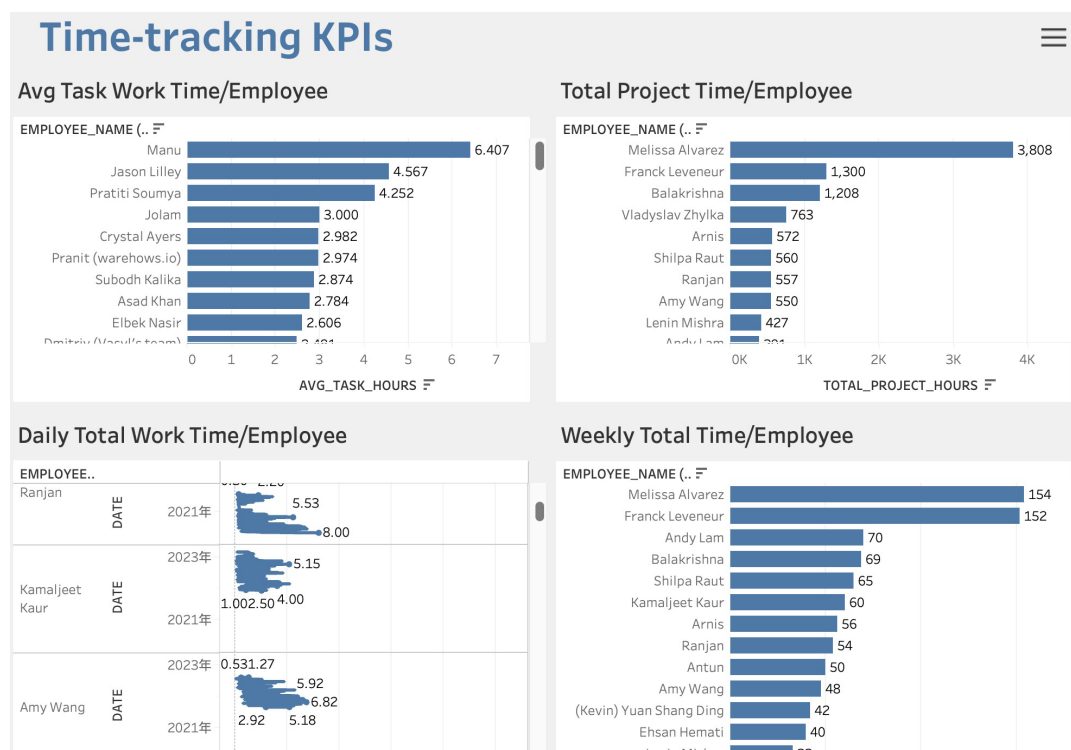
```
//Categorize the tasks by specific words
CASE
  WHEN CLOCKIFY_TASK_NAME LIKE '%Communication%' THEN 'Communication'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Technical Meeting%' THEN 'Meeting'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Meeting%' THEN 'Meeting'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Infrastructure%' THEN 'Infrastructure'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Discovery%' THEN 'Discovery'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Data%' THEN 'Data'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Onboarding%' THEN 'Onboarding'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Documentation%' THEN 'Documentation'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Management%' THEN 'Management'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Connectivity%' THEN 'Connectivity'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Proposal%' THEN 'Proposal'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Implementation%' THEN 'Implementation'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Optimization%' THEN 'Optimization'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Research%' THEN 'Research'
  WHEN CLOCKIFY_TASK_NAME LIKE '%Fact & Dim%' THEN 'Fact & Dim'
```

The code then joins the tables to extract the necessary data. It joins the DIM_CLOCKIFY_TASK, FACT_EMPLOYEE_TIME_ENTRY, DIM_CLOCKIFY_PROJECT, and DIM_CLIENT tables using the relevant keys.

In summary, the SQL code is used to extract relevant columns from the Clockify data, categorize projects and tasks based on specific keywords, join the relevant tables, and filter the data to exclude null values.

Visualization and Insights:

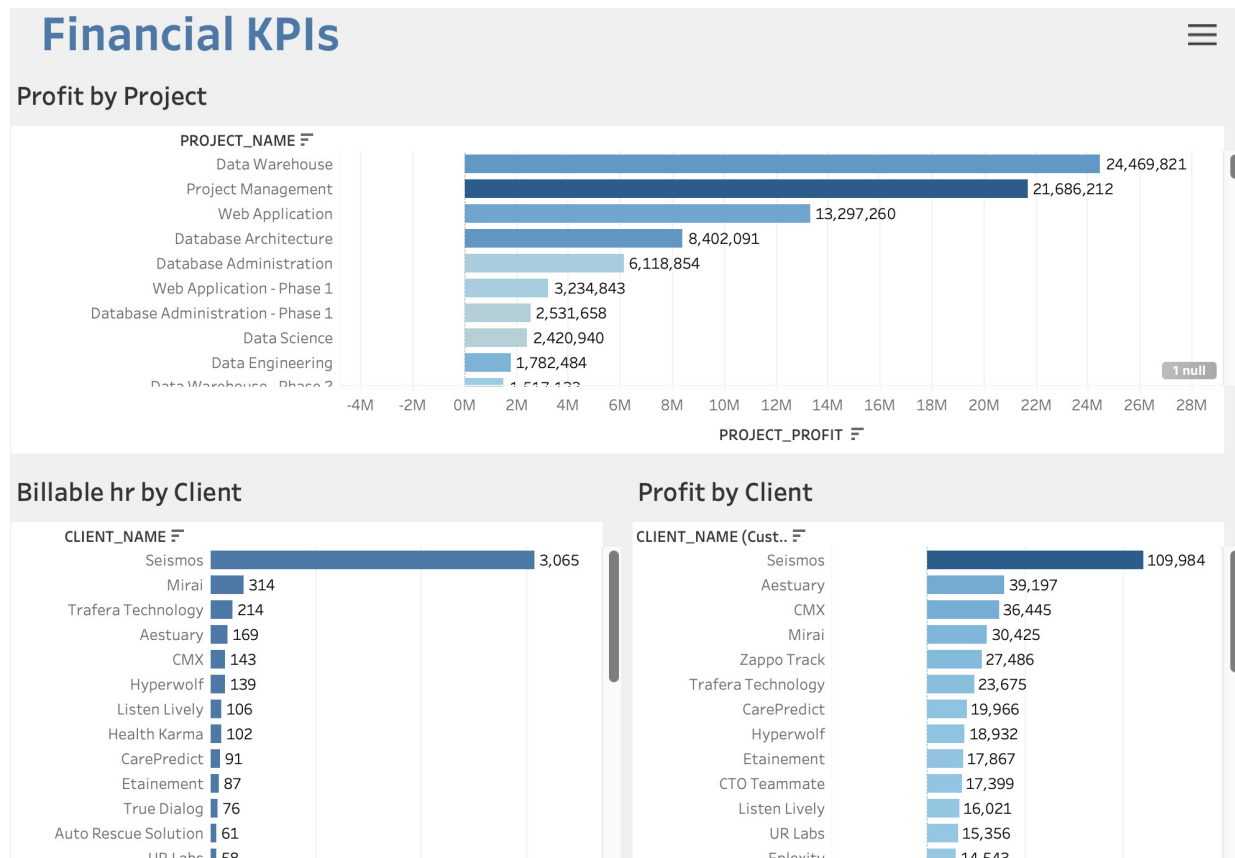
1. Time Tracking KPIs Dashboard



Tracking employee performance and workload is crucial for companies to ensure productivity and efficiency. The Time-tracking dashboard provides valuable insights into employee performance, and their productivity.

- I. **Top performing employees:** From the Time-tracking dashboard, we can see that Melissa Alvarez, Franck Leveneur, and Balakrishna are among the top 5 employees in both total project time/employee and weekly project time/employee table. This suggests their hard-workingness and ability to complete tasks within the allotted time frame.
- II. **Employee productivity analysis:** On the other hand, employee Manu has an average task work time significantly higher than the second-highest employee's. While this could imply his working capability, more information is needed to make a solid conclusion. Therefore, Clockify should consider conducting further analysis to identify the underlying factors causing this trend.

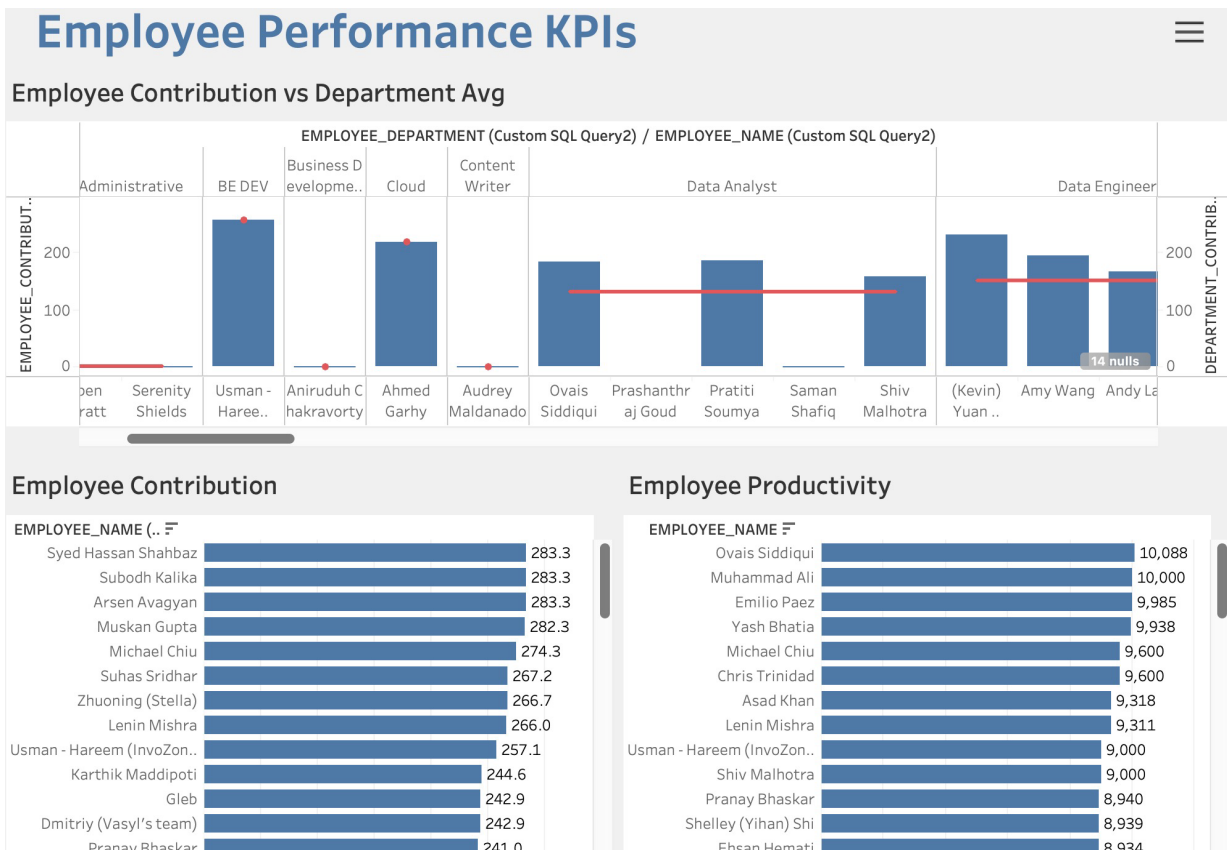
2. Financial KPIs Dashboard



Understanding financial performance is essential for any business to make informed decisions and plan for the future. The Financials dashboard provides a comprehensive view of the company's revenue, billable hours, and profitability by project type and client.

- I. **Revenue by project type:** According to the Financials dashboard, Data Warehouse projects and Project Management brought the most revenue for the company. This provides insights on the types of projects that Clockify has considerable audiences for and the ones they should focus on to maximize their revenue.
- II. **Top clients by billable hours and profit:** Furthermore, the top 5 clients in both the billable hour and profit tables are Seismos, Aestuary, and Mirai. This information is crucial for maintaining Clockify's relationships with these clients and identifying potential areas for growth.

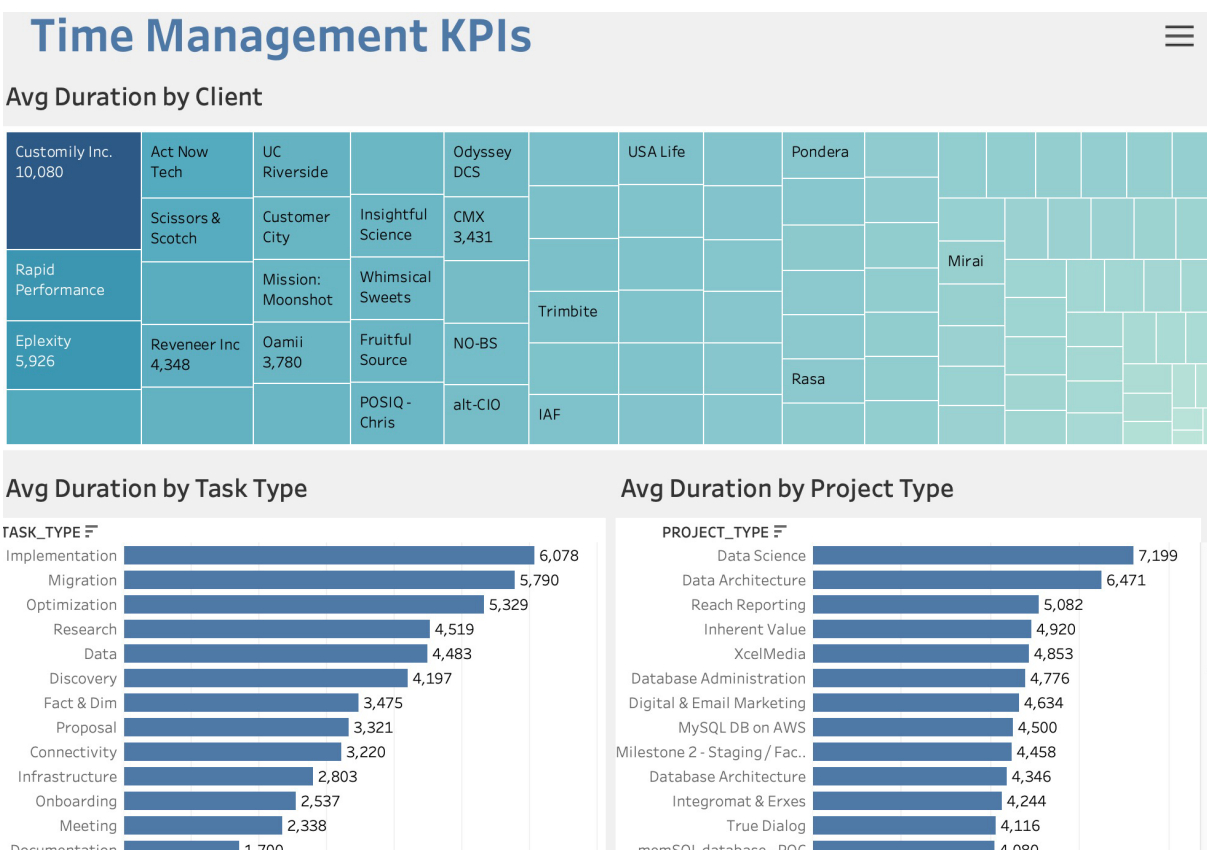
3. Employee Performance KPIs



Employee productivity and performance are critical factors for any organization's success. The Employee Performance dashboard offers valuable insights into employee performance, contribution, and cost-effectiveness.

- I. **Employee productivity and cost:** From the Employee Performance dashboard, Clockify should focus on the top and bottom employees from the Employee Productivity table. Additionally, they should conduct a cost analysis to determine whether the high productivity translates into higher profits or if it is cost-ineffective for the company.
- II. **Departmental performance:** Moreover, Clockify should be aware of the department or individual with significantly low contributions as they might be cost-ineffective to the company's operations. Therefore, conducting a performance analysis on a departmental and individual basis can help the company identify areas for improvement and optimize its resources.

4. Time Management



Finally, the Time Management visualization has revealed the following noteworthy findings:

- I. **Task duration:** The average duration of implementation tasks is the longest, followed by migration and optimization tasks. This suggests that these tasks may be more complex or require more resources to complete. In contrast, Communication, Management, and Documentation are tasks that have the shortest average duration, which may indicate that they are relatively straightforward.
- II. **Project duration:** Among the different project types, data science projects have the longest average duration, followed by data architecture and reach reporting projects. This could be due to the nature of these projects, which may involve more complex data analysis or modeling. In contrast, projects such as Meeting, Optimization, and Discovery have the shortest average duration, which may suggest that they are simpler or more routine in nature.
- III. **Client spending:** Customily Inc is the top spending client, followed by Rapid Performance, Eplexity, and Scissors&Scotch. This could indicate that these clients have a higher demand for the services offered by the company or require more extensive projects.

- IV. **Data-related projects:** Eplexity spends the most on data-related projects, followed by Trimbite, Reveneer, and Mirai. This suggests that these clients may have a greater need for data-related services or that the company has specialized expertise in this area.
- V. **Task allocation:** Meeting, Data, and Discovery tasks are the most commonly allocated tasks among clients. This could indicate that these tasks are essential components of many projects or that clients require a significant amount of time for communication and data-related activities. In contrast, implementation, migration, research, and fact & dim tasks are less commonly allocated, which may suggest that they are more specialized or require a narrower set of skills.

Project Challenges:

This was a challenging project that we collectively tackled. The open-endedness and complexity of the project brought us difficulties, but they also allowed us to put together a more sophisticated solution. We made use of multiple tools, references, and data sources. Throughout the process, there were a few challenges that we encountered, and we would like to discuss them in two aspects.

On the technical side, data reliability and data integrity is our top problem. Due to man-made errors, recorded data itself may not perfectly reflect the actual situation. For instance, in our dashboard, we have some data points showing an employee has more than 24 hours of daily working time. As we do not know what the “authentic” data looks like, we might give biased/inaccurate interpretations. We believe this is critical for a project like this; corporate executives base their decisions on data for the most part, and therefore the least we should do is to ensure that the data we collect is authentic and ethical, and that the metrics we provide are unbiased and straightforward. This would take a collective effort across multiple teams. We had limited transparency into certain parts of the data lifecycle for this project, but in the future, we would definitely like to work cross-functionally to ensure our data is accurate and authentic.

On the other side, we had some difficulties “understanding” the data. The existing documentation was helpful for us to get started, but it took some time for us to fully understand every single data table and attribute. Some of the namings of columns were not entirely intuitive. We had to look into the data and sometimes join across tables to figure out the meaning of a certain column, or whether the relationship between two entities is 1-to-1 or 1-to-many. For example, it took some time for us to figure out that $[\text{total revenue}] = [\text{the total amount of billable hours}] * [\text{hourly rate for that project}]$, and $[\text{total cost}] = \text{sum}([\text{billable hours}] * [\text{hourly rate for each employee}])$. These are fundamental pieces of information we had to sort out before designing the KPIs. The slightest misunderstanding could lead to incorrect metrics and KPIs, so we could not afford any “guesses”. Therefore, one of our key takeaways is that whenever someone designs a table or metric, that person should be responsible for keeping clear documentation of what each new variable means and how it is derived, as we believe it would have been much easier if there was more first-hand, detailed documentation for the data sources.

Regardless of the challenges, we were able to cruise through this project. Most of these issues could have been manually prevented, so it was truly valuable for us to pin down these challenges, reflect on how we would resolve them from the roots, and apply these lessons to our future data projects.