Task 1

There are a couple of ways that came to mind. But, I will mention here one technique.

1. First, as we need to learn a grouping structure, the problem is a clustering problem.
2. As the dataset is large, we can work with a churn of data. We can start working with a random sample of data(for example, 200 attributes and 2000 observations which randomly chosen). Then we can do analysis, clustering on it, and then make a result. Here, we can use the bootstrapping technique as our random sample is small. We can obtain a result.
3. We can do the same thing as the second step for the multiple random samples and do the same tasks which we did before. Similarly, we will interpret the result on these random samples.
4. In the final step, we will take the average of all of our results. Then we make a conclusion based on the average results of these random samples. We can use the ensemble voting technique on the sample dataset results to find an interpretable result for the entire dataset.

Task 2

N.B: I am good in MySQL database than PostgreSQL as I am confident with MySQL syntax. I know Postgre SQL as well. But, this database is not installed on my machine. So,I was not able to do more cross-check for Postgre SQl syntax. I am confused with highlighted part whether it will work or not.By this syntax I was trying to extract the year from the 'requested_at' attribute.I am giving my MySQL syntax as well, as it works fine.

Postgre SQL Syntax:

```
SELECT p.type, o.interest_rate, o.week_first_time_past_due, sum(o.expected_profit_per_week)

FROM open_loans o

JOIN product p ON p.id=o.id

JOIN clientdata c ON c.id=o.client_id

WHERE  date_part('year',TIMESTAMP,o.requested_at)= extract(year from current_date) AND
o.week_first_time_past_due<= EXTRACT('week' FROM current_date)

GROUP BY p.type, o.interest_rate, o.week_first_time_past_due
```

MySQL syntax:

Mysql:

```
SELECT p.type_, o.interest_rate, o.week_first_time_past_due, sum(o.expected_profit_per_week)

FROM open_loans o

JOIN product p ON p.id=o.id

JOIN clientdata c ON c.id=o.client_id

WHERE  DATE_FORMAT(o.requsted_at,'%Y')= DATE_FORMAT(curdate(),'%Y') AND
o.week_first_time_past_due<= WEEK(curdate())

GROUP BY p.type_, o.interest_rate, o.week_first_time_past_due
```

Task 3:

I am not writing separate report for this tasks. I tried explain everything what I did in the notebook file.

I have attached a notebook and html file for this tasks.