# **Insurance Data SQL Project**

This project is to showcase the insightful information one can extract from their membership details.

The project is based on a fictional insurance company based in Australia that have members in their company with various insurance covers supplied by various enterprise product suppliers. Based off my experience in the financial services industry, it is important to keep an overview of your client base for gaps in their financial planning, client relationship management and to know which product supplier is not only widely favoured but also cost-efficient for your clients’ needs

The project extrapolates the project work that was done during my Udemy SQL Data Analysis course and expands to find other insightful information for the DBA to pass on to the BI analyst hence the information is only expressed in SQL, the PowerBI dashboard will be incorporated in a separate project.

The database was provided for by the course instructor and imported into SSMS.

The objective of this project is to display the following skills:

* Understanding of SQL syntax and relational database operations including JOINS, CASE WHEN statements, Conditional queries and Aggregation functions.
* Understanding of Data analysis requirements and the relevant data that brings insight to the business’ operations.
* To conduct a feasibility study of the business’s operations

Glossary

Enterprise – This would be the product supplier that provides the insurance cover for our members.

DTH – Death cover; insurance that pays out in the event of the death of the member.

IP – Income Protector; insurance that pays out in the event a member cannot perform their daily work functions.

TPD – Total and Permanent Disability Cover; insurance that pays out in the event of the disablement of the member.

TIB – Terminal Illness Benefit cover; insurance that pays out in the event of the death of the member.

The first was to explore the tables to get a feel of the data we’re working with and what primary keys we might be working with, the subsequent relationships. We also want to have a look at whether there are any null values, missing values and/or incorrect data that needs to be manipulate or removed.

The queries that I performed are as follows:

1. Query that breaks down the membership demographics by age group and gender. We want to know what our client base looks like. As a financial planner, I have an ideal client profile and would like to know how much of my client base fits the criteria.

Our largest member group are males between the ages of 33 and 40, with the largest female group being in the same age range. For our business that can be good as the customer lifetime will be long, assuming life expectancy is 83 (according to Australian Institute of Health & Welfare, circa 2021)

1. Subsequently I want to have a look at the average annual salary of my members and their occupation to see how well they fit my ideal client profile. There are missing/incorrect entries in the employee status and I chose to classify all of these as inactive as it seemed to be the best logical fit (The annual\_salary was 25000 for all which seems to be how the DBA chose to tag these particular individuals with missing employee status information)

Our largest group is ‘Not Stated’, which we chose not to manipulate as this information can be corrected during a client review process or through a customer survey. The top 4 occupations in terms of member count are ‘Data Enablement PM, LAN Administrator, ETL Testing Analyst, Computer Scientist’ which tells us that our company has a strong focus on the information technology sector thus when we do product design of their covers, we can take this into consideration.

1. I want to query the geographical breakdown of our members to see which states they’re located. There are missing values in both the postal code field in the member table and the postal code table and this will cause a duplication when we run our join statements. As I’m not Australian, I chose not to replace the missing values with correct state information or correct postal code as it would be an extensive research task beyond the scope of this project and seeing as the postal code would be incorrect both the postal code table and the member table then the matching would still happen which is our objective.

We see that most of our members reside in the New South Wales state with 8 members residing outside of Australia.

**Note**: For the first 3 queries, I could’ve run a N-tile partition by statement to get the ranking of the top category however I want to represent this finding graphically on Power BI so that’ll be for the PowerBI project on the same dataset

1. Our analysis moves over to the member cover details in the MemberCover table of which we can see that there are missing values for the total\_death\_cover and total\_death\_premium fields which could be as a result of the member not having that cover in place. With that in mind I chose to replace the missing values with a 0 for quantitative calculation purposes.

I want to know what the average cover value and premiums collected for each cover category, and which one of our product suppliers has the most member count. The purpose of this is have an idea of which supplier is cost effective (by comparing the average cover amount and premiums per category) and which one of those enterprises is most favoured amongst our client base.

1. Next, I want to conduct a similar analysis of cover and premiums by Occupation. The reason why this is important is when we do a gap-analysis of our client’s risk planning we can identify the areas in which there is a gap in cover and focus our literacy and/or marketing campaigns towards these group of clients to address the gap.

During product design, this analysis will prove useful as we will be able to have a detailed breakdown of how the risk rating per occupation category will influence the overall premiums of the package and should we chose to have an industry specific product and/or an education linked product (i.e 3-year graduate’s cover or a Professionals package)

1. Now that we have a understanding of what our MemberCover details look like, I now want to understand what the claims history has been for our member base. I have queried the total claims paid and average age of the member that claimed, then grouped by Claim year, gender and ClaimType.
   1. The first analysis is the total claims paid and the average age. This important to note as the assumption of life expectancy is not true for the average client in our member base. The life expectancy in Australia is 83 but the average age of our DTH claims is 44 for both males and females which needs to be investigated further.
   2. The second analysis we conduct is the total number of claims and the claim amounts paid per year. This is to view if there are claim trends developing that we need to be aware of and to provide marketing with the statistics for our clients to be aware of when we inform them of the gaps in their cover.
2. The last query I want to conduct is a profitability study of our client base. I want to know if we’re making a profit on all the premiums we’re collecting. As I will be looking for all the premiums collected and not just by the members with claims, I need to create a view of the total premiums collected to perform correlated subqueries (using outer apply) for the claims paid vs premiums collect for each category.

We use the UNION statement to collate all the information we generated into one table as we want to analyse which product categories are profitable per underwriting year.

We see that the DTH covers are costing the company quite a lot which could be explained by the average age of the death claimants being significantly lower than the life expectancy. We either need to investigate whether there is fraud or if there is something wrong with our underwriting process.

Note: Queries 6 & 7 are not structured in the best way and could be improved upon however the objective of this project is to show an understanding of the SQL syntax and query structuring I did not think it was a great use of time to delve too deep into getting every detail correct but I am aware of where the improvements can be made

In all the queries, dynamic queries seemed to be the best option for finding null values in all my tables however I found it difficult to structure one and execute thus I left it out for the sake of simplicity.

During the editing of the query scripts and composing this text file, I considered erasing the text in the SQL script however I left the option of being able to read a summarized version of this project file for curios observers.

Hope you found my analysis insightful and pleasurable. You are more than welcome to contact me for further information or if you want to collaborate on a project

Email: [makalamabotjasfp@gmail.com](mailto:makalamabotjasfp@gmail.com)