

COMP8090 – Data and Visualisation for Business

ASSESSMENT TASK 1

Report on Database Design for ProApp

BY

Yuxiang Jeremy HONG - 46228241

2022 Session 1



MACQUARIE
University
SYDNEY • AUSTRALIA

Macquarie University
Sydney, Australia

Submission Date: 27 April 2022

Introduction

ProApp is an online and offline platform that serves to connect grassroots skilled people directly with the customer. It aims help tradespeople in Indonesia, who may not have any formal education or recognition of their trade skills and are being paid minimum rates. On the other hand, customers would have concerns and doubts about their personal security and skills of trades people when hiring someone they have never worked with before and no proof of skills.

ProApp aims to bridge this gap by empowering grassroot skilled people by providing the necessary police checks to address customers' security concerns and skills assessment and training in order to provide documentation for their skills in their trade.

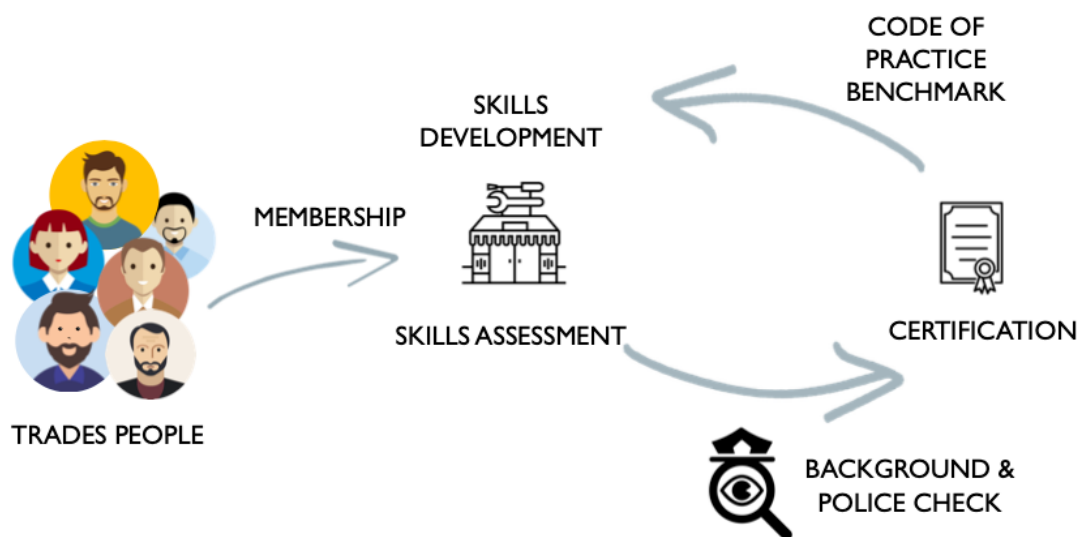


Figure 1: Onboarding Process of Trades People - Source: Business Prospectus

The business model of the ProApp earns revenue from customers through Task Quote fees, membership fees from Tradespeople and Suppliers, and transaction fees from completed jobs through the app.

THE PLATFORM

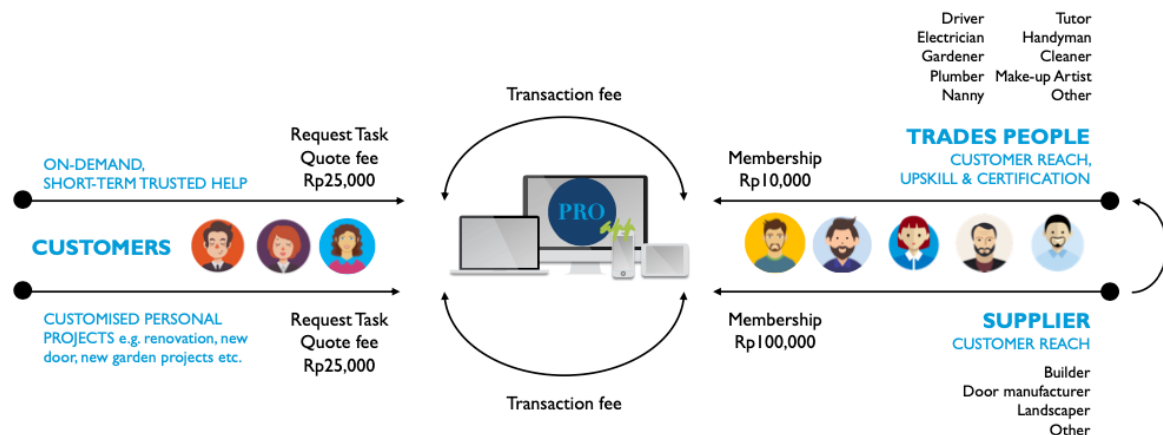


Figure 2: Business Model of ProApp - Sources: Business Prospectus

Objective

The objective of this assessment task is to:

- Build and implement a SQL database for ProApp
- Assist COO in investigating and gathering insights on how to minimize business overheads of managing newly registered users.

Database For ProApp

Assumptions

For the scope of this assessment task, the following assumptions have been made and used in the design of the database:

1. All app functions work perfectly, and database data is updated as intended.
2. Payment between tradesperson and supplier runs independently and perfectly outside of the app.
3. All issued tasks are accepted and executed by the tradesperson without issues.
4. Itemization of tasks and materials not required.
5. Australian Postcode and phone number structure used in dummy data for this assessment for ease of simulation.

Database Design Overview

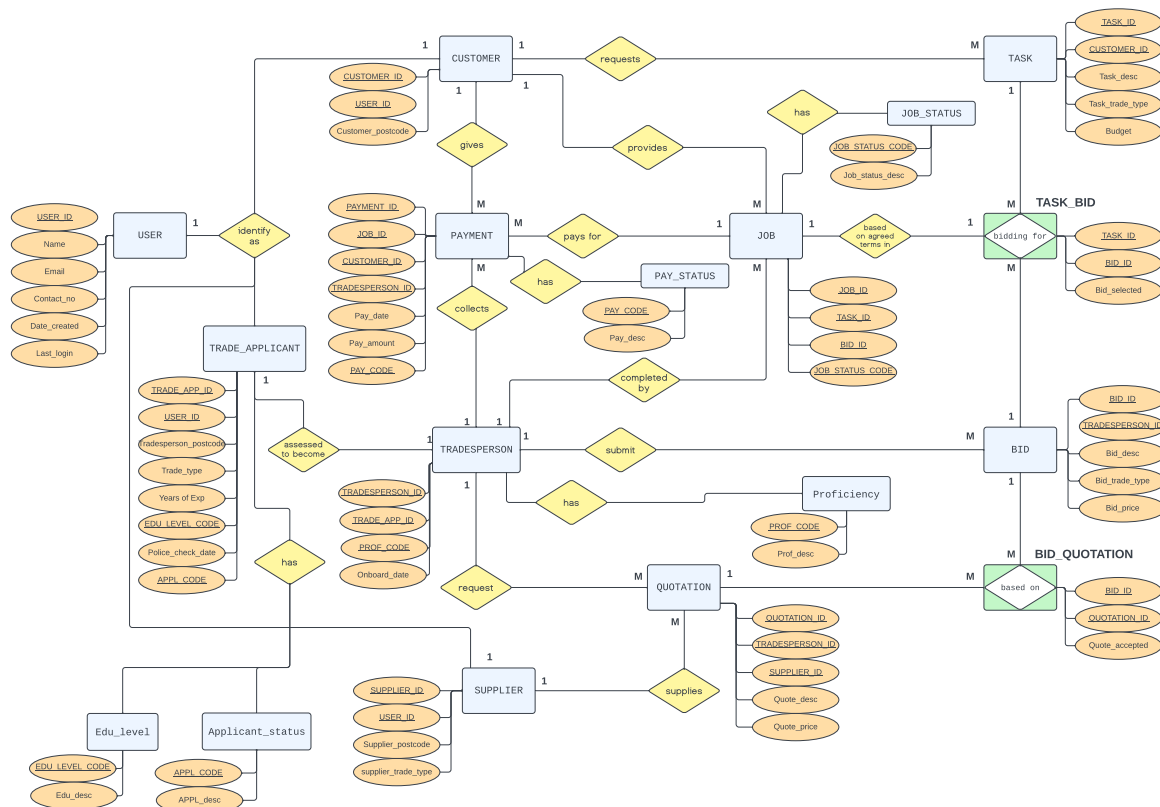
In considering the design of the database, we have to walk through the user experience and map out the entities involved. In doing so, we work out the tables that are required and how they relate to each other.

In an ideal experience for a user that we are aiming for is as follows:

- New users sign up through the app or website
 - They will choose to be Customer, Tradesperson or Supplier
- To become a Tradesperson,
 - users have to go through assessment and checks as a Trade Applicant. This is to ensure security and quality of work received by customers. This process includes:
 - Skill Assessment at a skill assessment centre
 - Police checks at the relevant government agency
 - Once the Trade Applicant has completed the checks and assessments, they are approved to become Tradesperson.
 - A Proficiency level of their trade would be assigned.
- In order to matchmake Jobs:
 - Customer will submit requests for Tasks for the services they require.
 - Tradespeople will submit bid for the Tasks.
 - If required, Tradespeople will request for quotation from suppliers.
- When a Bid for the Task is accepted, a Job is created and both Customers and Tradesperson will be able to communicate and provide the services as required.
- Payment can be made at any point of the Job based on the mutual agreement between Customer and Tradesperson.
 - Multiple payments can be made for a job: deposit, partial payment, full payment.

Entity Relationship Diagram (ERD)

With the above process in mind, an Entity Relationship Diagram (ERD) has been created to help visualise the tables required for the database and the relationships between tables. A higher resolution of the ERD is available in the Appendices.



Database Dictionary

Once the ERD is done, a data dictionary is created to clearly define the Primary and Foreign keys, the attributes, and the data types to be used for each attribute in each table.

TABLE NAME	ATTRIBUTE NAME	CONTENTS	TYPE	NULL Status	FORMAT	KEY	FK REFERENCED TABLE
USER	USER_ID	User ID	INT	NOT NULL	9999	PK	
	Name	User Name	VARCHAR(500)	NOT NULL	Xxxxx		
	Email	User Email	VARCHAR(500)	NOT NULL	Xxxxxxxxxxx		
	Contact_no	User Contact number	INT	NOT NULL	9999999999		
	Date_created	Date User account created	DATE	NOT NULL	YYYY-MM-DD		
	Last_login	Date Last login	DATE	NOT NULL	YYYY-MM-DD		
EDU_LEVEL	EDU_LEVEL_CODE	Education level in integer code	TINYINT	NOT NULL	9	PK	
	Edu_desc	Education Level description	VARCHAR(30)	NOT NULL	Xxxxxxx		
PROFICIENCY	PROF_CODE	Assessed Proficiency level in integer code	TINYINT	NOT NULL	9	PK	
	Prof_desc	Assessed Proficiency level description	VARCHAR(30)	NOT NULL	Xxxxxxx		
APPLICANT_STATUS	APPL_CODE	Status of Applicant in integer code	TINYINT	NOT NULL	9	PK	
	Appl_desc	Status of Applicant description	VARCHAR(30)	NOT NULL	Xxxxxxx		

PAY_STATUS	PAY_CODE	Status of payment for job in integer code	TINYINT	NOT NULL	9	PK	
	Pay_desc	Status of payment for job description	VARCHAR(30)	NOT NULL	Xxxxxxxx		
JOB_STATUS	JOB_STATUS_CODE	Status of Job progress in integer code	TINYINT	NOT NULL	9	PK	
	Job_status_desc	Status of Job progress description	VARCHAR(30)	NOT NULL	Xxxxxxxx		
CUSTOMER	CUSTOMER_ID	Customer ID	INT	NOT NULL	999	PK	
	USER_ID	User ID	INT	NOT NULL	9999	FK	USER
	Customer_postcode	Customer home postcode	INT	NOT NULL	9999		
SUPPLIER	Supplier_ID	Supplier ID	INT	NOT NULL	999	PK	
	USER_ID	User ID	INT	NOT NULL	9999	FK	USER
	Supplier_postcode	Supplier Post code	INT	NOT NULL	9999		
	Trade_type	Type of Trade of supplier	VARCHAR(20)	NOT NULL	Xxxxxxxxxxxx		
TRADE_APPLICANT	TRADE_APP_ID	Tradesperson Applicant ID	INT	NOT NULL	999	PK	
	USER_ID	User ID	INT	NOT NULL	9999	FK	USER
	tradesperson_postcode	Tradesperson Post Code	INT	NOT NULL	9999		
	trade_type	Type of Trade of Tradesperson	VARCHAR(20)	NOT NULL	Xxxxxxxxxxxx		
	experience	Tradesperson experience in years	INT(2)	NULL	99		
	EDU_LEVEL_CODE	Education level in integer code	TINYINT	NULL	9	FK	EDU_LEVEL
	police_check_date	Date of Police Check completed	DATE	NULL	YYYY-MM-DD		
	APPL_CODE	Status of Applicant in integer code	TINYINT	NOT NULL	9	FK	APPLICANT_STATUS
TRADEPERSON	TRADEPERSON_ID	Approved Tradesperson ID	INT	NOT NULL	999	PK	
	TRADE_APP_ID	Tradesperson Applicant ID during application	INT	NOT NULL	9999	FK	TRADE_APPLICANT
	PROF_CODE	Assessed Proficiency level in integer code	TINYINT	NOT NULL	9	FK	PROFICIENCY
	Onboard_date	Date of Approval of Tradesperson	DATE	NOT NULL	YYYY-MM-DD		
TASK	TASK_ID	Task ID	INT	NOT NULL	999	PK	
	CUSTOMER_ID	Customer ID of issued task	INT	NOT NULL	9999	FK	CUSTOMER
	Task_desc	Description of task	VARCHAR(5000)	NOT NULL	Xxxxxxxxxxxx		
	Task_trade_type	Type of trade of Task	VARCHAR(20)	NOT NULL	Xxxxxxxxxxxx		
	Budget	Budget in Rupiah	INT	NOT NULL	999999999999		
QUOTATION	QUOTATION_ID	Quotation ID	INT	NOT NULL	999	PK	
	TRADEPERSON_ID	Tradesperson ID requesting Quotation	INT	NOT NULL	9999	FK	TRADEPERSON
	SUPPLIER_ID	Supplier ID supplying Quotation	INT	NOT NULL	9999	FK	SUPPLIER
	Quote_desc	Description of Quotation	VARCHAR(5000)	NOT NULL	Xxxxxxxxxxxx		
	Quote_price	Quotation Price in Rupiah	INT	NOT NULL	999999999999		
BID	BID_ID	Bid ID	INT	NOT NULL	999	PK	
	TRADEPERSON_ID	Tradesperson ID submitting bid for task	INT	NOT NULL	9999	FK	TRADEPERSON
	Bid_desc	Description of Bid	VARCHAR(5000)	NOT NULL	Xxxxxxxxxxxx		
	Bid_trade_type	Type of Trade of bid	VARCHAR(20)	NOT NULL	Xxxxxxxxxxxx		
	Bid_price	Price of Bid for Task	INT	NOT NULL	999999999999		
BID_QUOTATION	BID_ID	Bid ID	INT	NOT NULL	999	FK	BID
	QUOTATION_ID	Quotation ID	INT	NOT NULL	999	FK	QUOTATION
	Quote_accpeted	Quote accepted for Bid for Task	BOOL	NULL	1		
TASK_BID	TASK_ID	Task ID	INT	NOT NULL	999	FK	TASK
	BID_ID	Bid ID	INT	NOT NULL	999	FK	BID
	Bid_selected	Selected Bid for Task	BOOL	NULL	1		
JOB	JOB_ID	Job ID generated after Bid accepted by Task	INT	NOT NULL	999	PK	
	TASK_ID	Task ID	INT	NOT NULL	999	FK	TASK_BID
	BID_ID	Bid ID of accepted Bid	INT	NOT NULL	999	FK	TASK_BID
	JOB_STATUS_CODE	Status of Job progress in integer code	TINYINT	NOT NULL	9	FK	JOB_STATUS
PAYMENT	PAYMENT_ID	Payment ID	INT	NOT NULL	999	PK	
	JOB_ID	Job ID of payment	INT	NOT NULL	999	FK	JOB
	CUSTOMER_ID	Customer ID making payment	INT	NOT NULL	999	FK	CUSTOMER
	TRADEPERSON_ID	Tradesperson ID receiving payment	INT	NOT NULL	999	FK	TRADEPERSON
	Pay_date	Date payment is made	DATE	NOT NULL	YYYY-MM-DD		
	Pay_amount	Amount paid in Rupiah	INT	NOT NULL	999999999999		
	PAY_CODE	Status of payment for job in integer code	TINYINT	NOT NULL	9	FK	PAY_STATUS

Business Case Queries

Formulating Queries from Business Case

In order to achieve the objective of the Chief Operating Officer to minimise the overheads of managing newly registered user. We have to query the database to extract the data we need to analyze and obtain insights that would help with decision-making.

From a business point of view, one of the first places to check in order to minimise overheads is to identify inefficiencies in the user onboarding process. The user-onboarding process for Customers and Suppliers are fairly simple. However, the process of onboarding Tradespersons are a bit more complex as the applicants are required to have their skills assessed, and their background checked out. This would be where we will be focusing to discover any potential for inefficiencies within the process.

The key business operational questions here would be:

- How long does it take to onboard a new tradesperson?
 - Who took the longest time to onboard?
 - Who took the shortest time to onboard?
 - Is there anything we can learn from the above two cases?
 - What is the average time taken to onboard a new tradesperson?
- Are our resources efficiently allocated?
 - Where should we locate skill assessment centres?
 - Which postcodes have the most applicants to be a tradesperson?
 - In each skill assessment centre, what trade skill assessment services have more demand?
- What number of applicants are in each stage of approval?
 - Where should we focus onboarding efforts more?
 - How does it look like for each trade?
- Are there enough tradespeople to meet the demands of customers?

Suggested Return Results

From the business questions, we will be able to generate queries in our database to generate data that we can analyze to obtain insights.

My main approach towards querying multiple questions is that the query for each business question is saved as a View and recalled as an output. This provides an ease to running the whole SQL file and generate all the recurring query outputs quickly. With each query saved as a view, they can be quickly recalled as well.

Some of the suggest query results for the above questions will be discussed here. The full query code is included in the query SQL file.

Creating Tradespeople Onboarding View:

```
DROP VIEW IF EXISTS Tradespeople_Onboarding;

CREATE VIEW Tradespeople_Onboarding AS
SELECT t1.TRADE_APP_ID, t1.USER_ID, trade_type, Date_created, police_check_date, onboard_date,
       -DATEDIFF(Date_created, police_check_date) as days_to_police_check,
       -DATEDIFF(Date_created, onboard_date) as days_to_onboard
FROM TRADE_APPLICANT as t1
INNER JOIN TRADEPERSON as t2
ON t1.TRADE_APP_ID = t2.TRADE_APP_ID
INNER JOIN USER as t3
ON t1.USER_ID = t3.USER_ID
;

# Create view of query
SELECT * FROM Tradespeople_Onboarding
ORDER BY days_to_onboard DESC, days_to_police_check DESC;
```

SQL Output:

	TRADE_APP_ID	USER_ID	trade_type	Date_created	police_check_da...	onboard_date	days_to_police_che...	days_to_onboard	
▶	5	9	Property	2021-04-09	2021-04-14	2021-05-15	5	36	
	1	5	Property	2021-04-05	2021-04-14	2021-05-10	9	35	
	7	11	Motor	2021-04-11	2021-04-18	2021-05-15	7	34	
	3	7	Motor	2021-04-07	2021-04-14	2021-05-07	7	30	
	2	6	Personal Services	2021-04-06	2021-04-14	2021-05-02	8	26	
	4	8	Motor	2021-04-08	2021-04-14	2021-05-04	6	26	

1a. Who took the longest time to onboard?

Query Code:

```
# Create view of query
SELECT * FROM Tradespeople_Onboarding
ORDER BY days_to_onboard DESC, days_to_police_check DESC;

# 1a Tradesperson that took the longest time to onboard
DROP VIEW IF EXISTS Q1a_longest_case;

CREATE VIEW Q1a_longest_case AS
SELECT * FROM Tradespeople_Onboarding
WHERE days_to_onboard = (SELECT MAX(days_to_onboard) FROM Tradespeople_Onboarding);

SELECT * FROM Q1a_longest_case;
```

SQL Output:

	TRADE_APP_ID	USER_ID	trade_ty...	Date_created	police_check_da...	onboard_date	days_to_police_che...	days_to_onboard	
▶	5	9	Property	2021-04-09	2021-04-14	2021-05-15	5	36	

1b. Who took the shortest time to onboard?

Query Code:

```
# 1b Tradesperson that took the shortest time to onboard
DROP VIEW IF EXISTS Q1b_fastest_case;

CREATE VIEW Q1b_fastest_case AS
SELECT * FROM Tradespeople_Onboarding
WHERE days_to_onboard = (SELECT MIN(days_to_onboard) FROM Tradespeople_Onboarding);

SELECT * FROM Q1b_fastest_case;
```

SQL Output:

	TRADE_APP_ID	USER_ID	trade_type	Date_created	police_check_da...	onboard_date	days_to_police_che...	days_to_onboard	
►	2	6	Personal Services	2021-04-06	2021-04-14	2021-05-02	8	26	
	4	8	Motor	2021-04-08	2021-04-14	2021-05-04	6	26	

1c. What is the average time taken to onboard a new tradesperson?

Query Code:

```
# 1c. Average amount of time to onboard a new tradesmen
DROP VIEW IF EXISTS Q1c_avg_onboard_time;

CREATE VIEW Q1c_avg_onboard_time AS
SELECT AVG(days_to_onboard) FROM Tradespeople_Onboarding;

SELECT * FROM Q1c_avg_onboard_time;
```

SQL Output:

	AVG(days_to_onboard)	
►	31.1667	

2a. Which postcodes have the most applicants to be a tradesperson?

Query Code:

```
#2. How many tradesperson assessment centers should we have and where?

# 2a. Which postcode has the most number of Tradeperson applicants? (Where to place assessment centers?)
DROP VIEW IF EXISTS Q2a_most_trade_appl_postcode;

CREATE VIEW Q2a_most_trade_appl_postcode AS
SELECT Tradesperson_postcode, COUNT(TRADE_APP_ID) as Tradie_count
FROM TRADE_APPLICANT
GROUP BY Tradesperson_postcode
ORDER BY COUNT(TRADE_APP_ID) DESC
;

SELECT * FROM Q2a_most_trade_appl_postcode;
```

SQL Output:

	Tradesperson_postcode	Tradie_count	
►	2155	6	
	2154	4	
	2153	3	

2b. In each skill assessment centre, what trade skill assessment services have more demand?

Query Code:

```
# 2b. In each post code area, what are proportions of types of tradepeople?
# (What type of assessment services are required more at each center?)
DROP VIEW IF EXISTS Q2b_trade_appl_prop;

CREATE VIEW Q2b_trade_appl_prop AS
SELECT Trade_type, Tradesperson_postcode, COUNT(TRADE_APP_ID) as Tradie_count
FROM TRADE_APPLICANT
GROUP BY Trade_type, Tradesperson_postcode
ORDER BY Trade_type DESC, Tradesperson_postcode ASC, COUNT(TRADE_APP_ID) DESC
;

SELECT * FROM Q2b_trade_appl_prop;
```

SQL Output:

	Trade_type	Tradesperson_postco...	Tradie_count	
►	Property	2153	3	
■	Personal Services	2155	3	
■	Motor	2154	4	
■	Motor	2155	3	

3a. What number of applicants are in each stage of approval?

Query Code:

```
# 3a. What number of applicants are in each stage of approval?
# (Where should we focus onboarding efforts more?)
# (Rejected, Not yet started, in progress, in training, approved)
DROP VIEW IF EXISTS Q3_onboard_stages;

CREATE VIEW Q3_onboard_stages AS
SELECT Appl_desc, COUNT(TRADE_APP_ID)
FROM TRADE_APPLICANT as t1
INNER JOIN APPLICANT_STATUS as t2
ON t1.APPL_CODE = t2.APPL_CODE
GROUP BY t2.APPL_CODE
;

SELECT * FROM Q3_onboard_stages;
```

SQL Output

	Appl_desc	COUNT(TRADE_APP_ID)	
►	Rejected	1	
■	Not Yet Started	1	
■	In Progress	4	
■	In training	1	
■	Approved	6	

3b. How does it look like for each trade?

Query Code:

```
#3b by Trade
DROP VIEW IF EXISTS Q3_onboard_stages_bytrade;

CREATE VIEW Q3_onboard_stages_bytrade AS
SELECT Trade_Type, Appl_desc, COUNT(TRADE_APP_ID) as Tradie_count
FROM TRADE_APPLICANT as t1
LEFT JOIN APPLICANT_STATUS as t2
ON t1.APPL_CODE = t2.APPL_CODE
GROUP BY t2.APPL_CODE, Trade_Type
ORDER BY Trade_Type DESC, t2.APPL_CODE ASC
;

SELECT * FROM Q3_onboard_stages_bytrade;
```

SQL Output:

	Trade_Type	Appl_desc	Tradie_count	
►	Property	In Progress	1	
■	Property	Approved	2	
■	Personal Services	In Progress	2	
■	Personal Services	Approved	1	
■	Motor	Rejected	1	
■	Motor	Not Yet Started	1	
■	Motor	In Progress	1	
■	Motor	In training	1	
■	Motor	Approved	3	

4. Are there enough tradespeople to meet the demands of customers?

Query Code:

```
# 4. Are there enough tradies to meet customer needs?
# (What are the proportion of Tradies to Customers?)
DROP VIEW IF EXISTS Tradies_postcode_count;
DROP VIEW IF EXISTS Customer_postcode_count;

# create tradie count view
CREATE VIEW Tradies_postcode_count AS
SELECT Tradesperson_postcode, COUNT(TRADEPERSON_ID) as Tradie_count
FROM TRADEPERSON as t1
INNER JOIN TRADE_APPLICANT as t2
ON t1.TRADE_APP_ID = t2.TRADE_APP_ID
GROUP BY Tradesperson_postcode
;

# create customer count view
CREATE VIEW Customer_postcode_count AS
SELECT Customer_postcode, COUNT(CUSTOMER_ID) as Customer_count
FROM CUSTOMER
GROUP BY Customer_postcode
;

# join both both views into a single output
DROP VIEW IF EXISTS Q4_Tradie_to_Cust;

CREATE VIEW Q4_Tradie_to_Cust AS
SELECT Tradesperson_postcode as Postcode, Tradie_count, Customer_count
FROM Tradies_postcode_count as v1
INNER JOIN Customer_postcode_count as v2
on v1.Tradesperson_postcode = v2.Customer_postcode
ORDER BY Postcode ASC
;

SELECT * FROM Q4_Tradie_to_Cust;
```

SQL Output:

	Postcode	Tradie_count	Customer_cou...	
▶	2153	2	1	
	2154	1	1	
	2155	3	2	

Conclusion

With a well-designed database, as data analyst, we will be able to collect and extract data from users to generate insights that can steer business decisions and improve user experience and generate more business revenue.

More development could be made to improve the ProApp database and make it more robust for all the different scenarios in real life user interaction between customers, tradespersons, and suppliers.

We are able to meet the objective of the COO to minimise overheads for managing new users by gathering insights from the data which includes, highlighting best and worst cases in the onboarding process for further review, identify stages of the onboarding process that may be lengthy and inefficient. Furthermore, we are able to establish queries to identify best locations to setup Skill Assessment Centres with the appropriate services to meet the demands of tradesperson applicants and minimise unnecessary overheads in the centres by allocating the right amount of resources to meet demand.

APPENDICES

