



STORED PROCEDURE QUERY

Team- 1



Team Member Details:

NAME	EMAIL ID
TARINEE	tarinee@s.amity.edu
PRASHANT KASHYAP	prashantkashyap5579@gmail.com
VARUN JOSE MADANU	varunjosemadanu@gmail.com
MUSTAFA OUN	wnvyvcgc@gmail.com
ANKIT RAJ	226301029@gkv.ac.in
ANCHAL KEDIA	anchalkedia45@gmail.com
VAISHNAVI GATTAGONI	gattagonivaishnavi@gmail.com
USWA SHARIQ	uswashariq@gmail.com

Report on Stored Procedure like Query on master table

1. Data Cleaning and Null Handling

a. Check for Null Values

```
-- Check for NULL in cd_user_id
SELECT * FROM "Master_Table" WHERE "cd_user_id" IS NULL;

-- Check for NULL in email
SELECT * FROM "Master_Table" WHERE "email" IS NULL;

-- Check for NULLs or unwanted values in gender
SELECT * FROM "MasterData"
WHERE gender IS NULL OR gender = 'Don%27t want to specify' OR gender = 'NULL';
```

b. Clean Gender Field

```
-- Convert unwanted gender values to SQL NULL
UPDATE "Master_Table" SET gender = NULL WHERE gender = 'Don%27t want to specify';
UPDATE "Master_Table" SET gender = NULL WHERE gender = 'NULL';
UPDATE "Master_Table" SET gender = NULL WHERE gender = '[null]';

-- Replace NULL with user-friendly label
UPDATE "Master_Table" SET gender = 'Didn't Mention' WHERE gender = 'NULL';
```

c. Clean Birthdate Field

```
-- Check for NULLs and 'NULL'
SELECT * FROM "Master_Table" WHERE "birthdate" IS NULL;
SELECT * FROM "Master_Table" WHERE "birthdate" = 'NULL';

-- Replace 'NULL' with 'Didn't Mention'
UPDATE "Master_Table" SET "birthdate" = 'Didn't Mention' WHERE "birthdate" = 'NULL';
```

d. Clean City and State

```
-- Check for 'NULL' in city and state
SELECT * FROM "Master_Table" WHERE "city" = 'NULL' OR "state" = 'NULL';

-- Replace with 'Didn't Mention'
UPDATE "Master_Table"
SET
    "city" = CASE WHEN "city" = 'NULL' THEN 'Didn't Mention' ELSE "city" END,
    "state" = CASE WHEN "state" = 'NULL' THEN 'Didn't Mention' ELSE "state" END
WHERE "city" = 'NULL' OR "state" = 'NULL';
```

e. Clean Country

```
-- Check for 'NULL' in country
SELECT * FROM "Master_Table" WHERE "country" = 'NULL';

-- Replace with 'Didn't Mention'
UPDATE "Master_Table" SET "country" = 'Didn't Mention' WHERE "country" = 'NULL';
```

f. Clean Education Fields

```
-- Check for 'NULL' in major, institution, and degree
SELECT * FROM "Master_Table"
WHERE "major" = 'NULL' OR "degree" = 'NULL' OR "institution" = 'NULL';

-- Replace with 'Didn't Mention'
UPDATE "Master_Table"
SET
    "major" = CASE WHEN "major" = 'NULL' THEN 'Didn't Mention' ELSE "major" END,
    "degree" = CASE WHEN "degree" = 'NULL' THEN 'Didn't Mention' ELSE "degree" END,
    "institution" = CASE WHEN "institution" = 'NULL' THEN 'Didn't Mention' ELSE
"institution" END
WHERE "major" = 'NULL' OR "degree" = 'NULL' OR "institution" = 'NULL';
```

2. Data Type Transformation

a. Convert Text to Date Format

```
ALTER TABLE "Master_Table"
ALTER COLUMN "UserCreateDate" TYPE DATE
USING "UserCreateDate"::timestamp::date;

ALTER TABLE "Master_Table"
ALTER COLUMN "UserLastModifiedDate" TYPE DATE
USING "UserLastModifiedDate"::timestamp::date;
```

b. Convert Apply Date

```
UPDATE "Master_Table"
SET "apply_date" = NULL
WHERE "apply_date" = 'NULL';

ALTER TABLE "Master_Table"
ALTER COLUMN "apply_date" TYPE DATE
USING "apply_date"::timestamp::date;
```

c. Convert Start and End Dates (Epoch to Timestamp to Date)

```
ALTER TABLE "Master_Table"
ALTER COLUMN "start_date" TYPE TIMESTAMP
USING TO_TIMESTAMP(CAST("start_date"::NUMERIC AS BIGINT) / 1000);

ALTER TABLE "Master_Table"
ALTER COLUMN "start_date" TYPE DATE
USING "start_date"::DATE;

ALTER TABLE "Master_Table"
ALTER COLUMN "end_date" TYPE TIMESTAMP
USING TO_TIMESTAMP(CAST("end_date"::NUMERIC AS BIGINT) / 1000);

ALTER TABLE "Master_Table"
ALTER COLUMN "end_date" TYPE DATE
USING "end_date"::DATE;
```

d. Convert Size to Numeric

```
ALTER TABLE "Master_Table"
ALTER COLUMN "size" TYPE NUMERIC
USING "size"::NUMERIC;
```

3. Column Removal

```
ALTER TABLE "Master_Table" DROP COLUMN "zip";
ALTER TABLE "Master_Table" DROP COLUMN "tracking_questions";
ALTER TABLE "Master_Table" DROP COLUMN "cohort_id";
```

4. Enrichment and Derived Fields

a. Birthdate Cleaning and Generation Classification

```
ALTER TABLE "Master_Table" ADD COLUMN "birthdate_clean" DATE;
ALTER TABLE "Master_Table" ADD COLUMN "generation_name" TEXT;

UPDATE "Master_Table"
SET "birthdate_clean" = TO_DATE("birthdate", 'DD-MM-YYYY')
WHERE "birthdate" ~ '^\d{2}-\d{2}-\d{4}$';

UPDATE "Master_Table"
SET "generation_name" = CASE
    WHEN EXTRACT(YEAR FROM "birthdate_clean") BETWEEN 1997 AND 2012 THEN 'Gen Z'
    WHEN EXTRACT(YEAR FROM "birthdate_clean") BETWEEN 1981 AND 1996 THEN 'Millennial'
    WHEN EXTRACT(YEAR FROM "birthdate_clean") BETWEEN 1965 AND 1980 THEN 'Gen X'
    WHEN EXTRACT(YEAR FROM "birthdate_clean") BETWEEN 1946 AND 1964 THEN 'Baby Boomer'
    WHEN EXTRACT(YEAR FROM "birthdate_clean") < 1946 THEN 'Silent Generation'
    WHEN EXTRACT(YEAR FROM "birthdate_clean") > 2012 THEN 'Gen Alpha'
    ELSE 'Unknown'
END
WHERE "birthdate_clean" IS NOT NULL;

-- Drop original birthdate
ALTER TABLE "Master_Table" DROP COLUMN "birthdate";
```

b. User Activity Days

```
ALTER TABLE "Master_Table" ADD COLUMN "User_Activity_Portal" INTEGER;

UPDATE "Master_Table"
SET "User_Activity_Portal" = DATE("UserLastModifiedDate") - DATE("UserCreateDate")
WHERE "UserLastModifiedDate" IS NOT NULL AND "UserCreateDate" IS NOT NULL;
```

c. User Engagement Duration

```
ALTER TABLE "Master_Table" ADD COLUMN "User_engage_opp" INTEGER;

UPDATE "Master_Table"
SET "User_engage_opp" = DATE("start_date") - DATE("apply_date")
WHERE "start_date" IS NOT NULL AND "apply_date" IS NOT NULL;
```

d. Opportunity Completion Gap

```
ALTER TABLE "Master_Table" ADD COLUMN "User_comp_gap" INTEGER;

UPDATE "Master_Table"
SET "User_comp_gap" = DATE("end_date") - DATE("start_date")
WHERE "start_date" IS NOT NULL AND "end_date" IS NOT NULL;
```

Summary Table

OPERATION	DESCRIPTION
NULL CLEANING	Standardized null values and placeholders in various fields
TYPE CONVERSION	Applied on date fields, size, etc., for consistency
COLUMN REMOVAL	Dropped unnecessary columns like zip, cohort_id, etc.
FEATURE ENGINEERING	Added new fields: birthdate_clean, generation_name, and 3 duration columns
FINAL ENHANCEMENTS	Standardized formats and calculated user behavioral insights