# **Team Yelp Project - SQL Code**

Team members: Gabriel Ascencio Morales, Luke Watts, Taylor Wilson, Leonie Zhang

*-- Summary Statistics #1: Total number of users in user table*SELECT user\_id, count(\*)

FROM user;

*-- Summary Statistics #2: Total number of cities*

SELECT count(distinct city)

FROM business;  
  
*-- Summary Statistics #3: Total number of businesses in business table*

SELECT business\_id, count(\*)

FROM business;  
  
*-- Summary Statistics #4: Most reviewed business*  
SELECT business\_id, name, review\_count

FROM business  
GROUP BY name, review\_count

ORDER BY review\_count DESC

LIMIT 1;

*-- Summary Statistics #5: Average stars per business*

SELECT avg(stars) as avgStars  
FROM business;

-- *Summary Statistics #6: Average review count per user*

SELECT avg(review\_count)

FROM user;

-- *Summary Statistics #7: Average review count per business*

SELECT avg(review\_count)

FROM business;

– *Summary Statistics #8: Most reviewed category*

SELECT categories, review\_count

FROM business

GROUP BY categories

ORDER BY review\_count DESC;

-- *Summary Statistics #9: How many businesses are in each category*

SELECT categories, COUNT(name) AS NumBusinesses

FROM business

GROUP BY categories

ORDER BY NumBusinesses DESC;

-- *Summary Statistics #10: City with the most reviews*

SELECT city, sum(review\_count) as reviews

FROM business

GROUP BY City

ORDER BY Reviews DESC

LIMIT 5;

**Business Question**

-- *Creates table with total check-ins and assigns them to “Day\_check\_ins” or “Night\_check\_ins”*

CREATE TEMPORARY TABLE T1 AS

SELECT business\_id, sum("time.Thursday.18:00" + "time.Thursday.15:00" + "time.Thursday.13:00" + "time.Wednesday.11:00" + "time.Wednesday.13:00" + "time.Wednesday.14:00" + "time.Wednesday.17:00"

+ "time.Wednesday.18:00" + "time.Sunday.18:00" + "time.Sunday.16:00" + "time.Sunday.14:00" + "time.Sunday.17:00" + "time.Friday.16:00" + "time.Friday.14:00" + "time.Friday.18:00"

+ "time.Friday.10:00" + "time.Friday.17:00" + "time.Friday.15:00" + "time.Saturday.18:00" + "time.Saturday.10:00" + "time.Saturday.12:00" + "time.Saturday.13:00"

+ "time.Saturday.15:00" + "time.Saturday.16:00" + "time.Saturday.17:00" + "time.Saturday.14:00" + "time.Monday.12:00" + "time.Monday.11:00" + "time.Monday.17:00" + "time.Monday.14:00" + "time.Monday.15:00" + "time.Monday.18:00" + "time.Tuesday.18:00" + "time.Tuesday.12:00" + "time.Tuesday.13:00" + "time.Tuesday.16:00" + "time.Tuesday.17:00" + "time.Tuesday.15:00") AS Day\_check\_ins,

sum("time.Thursday.21:00" + "time.Thursday.1:00" + "time.Thursday.4:00" + "time.Thursday.2:00" + "time.Thursday.20:00"

+ "time.Thursday.22:00" + "time.Thursday.19:00" + "time.Thursday.23:00" + "time.Wednesday.6:00" + "time.Wednesday.2:00" + "time.Wednesday.0:00" + "time.Wednesday.1:00" + "time.Wednesday.21:00"

+ "time.Wednesday.19:00" + "time.Wednesday.20:00" + "time.Sunday.19:00" + "time.Sunday.23:00" + "time.Sunday.21:00" + "time.Sunday.20:00"

+ "time.Sunday.6:00" + "time.Sunday.0:00" + "time.Sunday.2:00" + "time.Sunday.3:00" + "time.Friday.23:00" + "time.Friday.21:00" + "time.Friday.19:00"

+ "time.Friday.22:00" + "time.Friday.3:00" + "time.Friday.0:00" + "time.Saturday.21:00" + "time.Saturday.23:00" + "time.Saturday.2:00" + "time.Saturday.0:00" + "time.Saturday.1:00" + "time.Monday.4:00" + "time.Monday.23:00" + "time.Monday.20:00" + "time.Monday.19:00" + "time.Tuesday.4:00" + "time.Tuesday.21:00" + "time.Tuesday.20:00" + "time.Tuesday.23:00") AS Night\_check\_ins

FROM checkin

GROUP BY business\_id;

*-- Creates table with total check-ins for businesses in Las Vegas and determines the most representative time-frame*

CREATE TEMPORARY TABLE T2 AS

SELECT business.business\_id, business.name, business.stars, business.city, business.review\_count, business.categories, T1.Day\_check\_ins, T1.Night\_check\_ins,

CASE

WHEN T1.Day\_check\_ins > T1.Night\_check\_ins THEN "Day"

WHEN T1.Day\_check\_ins < T1.Night\_check\_ins THEN "Night"

ELSE "Equally relevant"

END AS "Most\_representative\_timeframe"

FROM business, T1

WHERE business.business\_id = T1.business\_id

AND city LIKE '%Vega%';

*-- Adds the number of day and night check-ins in Las Vegas*

SELECT sum(Day\_check\_ins), sum(Night\_check\_ins)

FROM T2;

*-- Calculates the average star rating per relevant time-frame*

SELECT Most\_representative\_timeframe, avg(stars)

FROM T2

GROUP BY Most\_representative\_timeframe;

*-- Calculates the review count per time-frame*

SELECT Most\_representative\_timeframe, sum(review\_count)

FROM T2

GROUP BY Most\_representative\_timeframe;

*-- Obtains business categories per number of night check-ins in descending order*

SELECT Most\_representative\_timeframe, categories, Night\_check\_ins

FROM T2

WHERE Most\_representative\_timeframe = 'Night'

GROUP BY Most\_representative\_timeframe, categories

ORDER BY Night\_check\_ins DESC

LIMIT 5;