1. Выведите количество пустых значений по колонкам CRIM, ZN, INDUS, CHAS, NOX (название колонки, кол-во пустых значений)

SELECT

'CRIM' AS Column\_name

,COUNT(CRIM) AS Count\_of\_nulls

FROM boston

WHERE CRIM IS NULL

UNION SELECT

'ZN' AS Column\_name

,COUNT(ZN) AS Count\_of\_nulls

FROM boston

WHERE ZN IS NULL

UNION SELECT

'INDUS' AS Column\_name

,COUNT(INDUS) AS Count\_of\_nulls

FROM boston

WHERE INDUS IS NULL

UNION SELECT

'CHAS' AS Column\_name

,COUNT(CHAS) AS Count\_of\_nulls

FROM boston

WHERE CHAS IS NULL

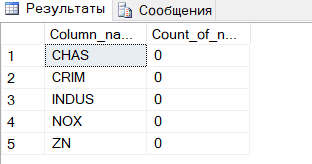
UNION SELECT

'NOX' AS Column\_name

,COUNT(NOX) AS Count\_of\_nulls

FROM boston

WHERE NOX IS NULL



2. Выведите количество уникальных значений по колонокам CRIM, ZN, INDUS, CHAS, NOX (название колонки, кол-во уникальных значений).

SELECT

'CRIM' AS Column\_name

,COUNT(DISTINCT CRIM) AS Unique\_values\_count

FROM boston

UNION SELECT

'ZN' AS Column\_name

,COUNT (DISTINCT ZN) AS Unique\_values\_count

FROM boston

UNION SELECT

'INDUS' AS Column\_name

,COUNT (DISTINCT INDUS) AS Unique\_values\_count

FROM boston

UNION SELECT

'CHAS' AS Column\_name

,COUNT (DISTINCT CHAS) AS Unique\_values\_count

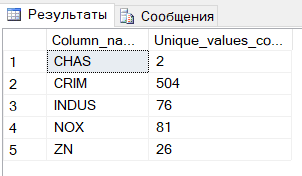
FROM boston

UNION SELECT

'NOX' AS Column\_name

,COUNT (DISTINCT NOX) AS Unique\_values\_count

FROM boston



3. Выведите колонки, у которых медиана равна минимальному значению (название колонки) выбирая из CRIM, ZN, INDUS, CHAS, NOX. Напишите какой вывод можно сделать по данным в этих колонках

WITH t AS (

SELECT

'CRIM' AS Column\_name

,(

(SELECT MAX(CRIM) FROM

(SELECT TOP 50 PERCENT CRIM FROM boston ORDER BY CRIM) AS BottomHalf)

+

(SELECT MIN(CRIM) FROM

(SELECT TOP 50 PERCENT CRIM FROM boston ORDER BY CRIM DESC) AS TopHalf)

) / 2 AS Median

,MIN(CRIM) AS MinValue

FROM boston

UNION SELECT

'ZN' AS Column\_name

,(

(SELECT MAX(ZN) FROM

(SELECT TOP 50 PERCENT ZN FROM boston ORDER BY ZN) AS BottomHalf)

+

(SELECT MIN(ZN) FROM

(SELECT TOP 50 PERCENT ZN FROM boston ORDER BY ZN DESC) AS TopHalf)

) / 2 AS Median

,MIN(ZN) AS MinValue

FROM boston

UNION SELECT

'INDUS' AS Column\_name

,(

(SELECT MAX(INDUS) FROM

(SELECT TOP 50 PERCENT INDUS FROM boston ORDER BY INDUS) AS BottomHalf)

+

(SELECT MIN(INDUS) FROM

(SELECT TOP 50 PERCENT INDUS FROM boston ORDER BY INDUS DESC) AS TopHalf)

) / 2 AS Median

,MIN(INDUS) AS MinValue

FROM boston

UNION SELECT

'CHAS' AS Column\_name

,(

(SELECT MAX(CHAS) FROM

(SELECT TOP 50 PERCENT CHAS FROM boston ORDER BY CHAS) AS BottomHalf)

+

(SELECT MIN(CHAS) FROM

(SELECT TOP 50 PERCENT CHAS FROM boston ORDER BY CHAS DESC) AS TopHalf)

) / 2 AS Median

,MIN(CHAS) AS MinValue

FROM boston

UNION SELECT

'NOX' AS Column\_name

,(

(SELECT MAX(NOX) FROM

(SELECT TOP 50 PERCENT NOX FROM boston ORDER BY NOX) AS BottomHalf)

+

(SELECT MIN(NOX) FROM

(SELECT TOP 50 PERCENT NOX FROM boston ORDER BY NOX DESC) AS TopHalf)

) / 2 AS Median

,MIN(NOX) AS MinValue

FROM boston

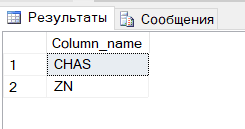
)

SELECT

Column\_name

FROM t

WHERE Median = MinValue



Вывод: Колонка CHAS принимает бинарные значения 0 или 1, поэтому медиана будет равна или 0 или 1

4. Выведите разницу между среднем количеством комнат(RM) в домах с самой дорогой стоимостью(MEDV) и 25 самыми дешевыми домами. Аналогично по 50, 100, 200, 300 самыми дешевыми домами. (кол-во домов(25,50,100,200,300), среднее кол-во комнат в них, среднее кол-во комнат в самых дорогих, разница). Напишите влияет ли кол-во комнат на стоимость и как сильно

GO

CREATE FUNCTION StatsByFlats (@FlatsCount INT )

RETURNS TABLE

AS RETURN (

WITH t1 AS (

SELECT TOP (@FlatsCount)

RM

FROM boston

ORDER BY CAST(MEDV AS FLOAT) DESC

),

t2 AS (

SELECT TOP (@FlatsCount)

RM

FROM boston

ORDER BY CAST(MEDV AS FLOAT) ASC

)

SELECT

CEILING(AVG(CAST(t1.RM AS FLOAT))) AS MaxPriceAVG

,CEILING (AVG(CAST(t2.RM AS FLOAT))) AS MinPriceAVG

,CEILING(AVG(CAST(t1.RM AS FLOAT)) )- CEILING(AVG(CAST(t2.RM AS FLOAT))) AS Diff

FROM t1

CROSS JOIN t2

)

SELECT

'25' AS FlatsCount

,MaxPriceAVG

,MinPriceAVG

,Diff

FROM StatsByFlats(25)

UNION SELECT

'50' AS FlatsCount

,MaxPriceAVG

,MinPriceAVG

,Diff

FROM StatsByFlats(50)

UNION SELECT

'100' AS FlatsCount

,MaxPriceAVG

,MinPriceAVG

,Diff

FROM StatsByFlats(100)

UNION SELECT

'200' AS FlatsCount

,MaxPriceAVG

,MinPriceAVG

,Diff

FROM StatsByFlats(200)

UNION SELECT

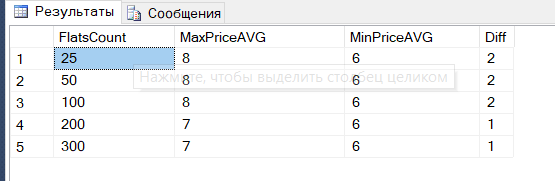
'300' AS FlatsCount

,MaxPriceAVG

,MinPriceAVG

,Diff

FROM StatsByFlats(300)



Вывод: Количество комнат влияет на стоимость квартир (с увеличением количества комнат растет стоимость квартир)

5. Выведите ранги значений колонки LSTAT(процент населения с более низким статусом) в домах с самой дорогой стоимостью (значение LSTAT, стоимость, ранг). Напишите какой вывод можно сделать по этим данным

SELECT \*

FROM (

SELECT

LSTAT

,MEDV,

RANK() OVER (ORDER BY LSTAT ASC) AS RANG FROM boston

) AS m

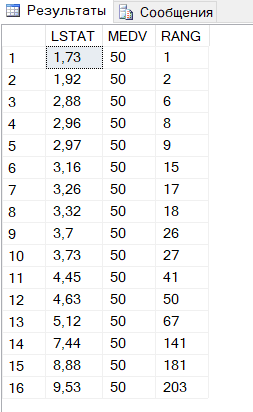
WHERE MEDV = (

SELECT

MAX(MEDV)

FROM boston

)



Вывод: Между данными прослеживается обратная корреляция – в районах, где продаются дорогие дома процент населения низкого статуса минимален.

6. Выведите среднюю стоимость домов граничащих с рекой(CHAS) и нет (граничит/не граничит, стоимость)

SELECT

(CASE

WHEN CHAS = 0

THEN 'Граничит'

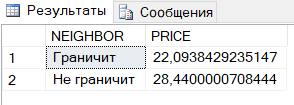
ELSE 'Не граничит'

END) AS NEIGHBOR,

AVG(MEDV) as PRICE

FROM boston

GROUP BY CHAS



7. Выведите все колонки, у которых среднее значение выше, когда дом граничит с рекой (название колонки) выбирая из CRIM, ZN, INDUS, CHAS, NOX. Напишите какой вывод можно сделать по этим данным.

WITH t AS (

SELECT

'CRIM' AS Column\_name

,(SELECT AVG(CRIM) FROM boston WHERE CHAS=1) AS NearRiver

,(SELECT AVG(CRIM) FROM boston WHERE CHAS=0) AS NotNearRiver

FROM boston

UNION SELECT

'ZN' AS Column\_name

,(SELECT AVG(ZN) FROM boston WHERE CHAS=1) AS NearRiver

,(SELECT AVG(ZN) FROM boston WHERE CHAS=0) AS NotNearRiver

FROM boston

UNION SELECT

'INDUS' AS Column\_name

,(SELECT AVG(INDUS) FROM boston WHERE CHAS=1) AS NearRiver

,(SELECT avg(INDUS) FROM boston WHERE CHAS=0) AS NotNearRiver

FROM boston

UNION SELECT

'CHAS' AS Column\_name

,(SELECT AVG(CHAS) FROM boston WHERE CHAS=1) AS NearRiver

,(SELECT AVG(CHAS) FROM boston WHERE CHAS=0) AS NotNearRiver

FROM boston

UNION SELECT

'NOX' AS Column\_name

,(SELECT AVG(NOX) FROM boston WHERE CHAS=1) AS NearRiver

,(SELECT AVG(NOX) FROM boston WHERE CHAS=0) AS NotNearRiver

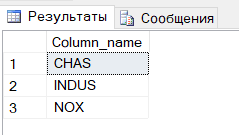
FROM boston

)

SELECT Column\_name

FROM t

WHERE NearRiver > NotNearRiver



Вывод: Доля неторгового бизнеса на город (INDUS) положительно коррелирует с концентрацией оксидов азота (NOX), т.е чем больше развит бизнес (что предполагает заводы и разрабатывающие предприятия), тем выше концентрация азота.

8. Выведите значения долей промышленной застройки(INDUS), концентрации оксидов азота(NOX) и по их перцентилям - 10, 20 ... 100 (перцентиль (10,20...100), значение INDUS, значение NOX). Напишите прослеживается между ними взаимосвязь.

SELECT

INDUS

,INDUS\_PERCENTILE

,NOX

,NOX\_PERCENTILE

FROM (

SELECT

INDUS

,NTILE(10) OVER (ORDER BY INDUS) \* 10 AS indus\_percentile

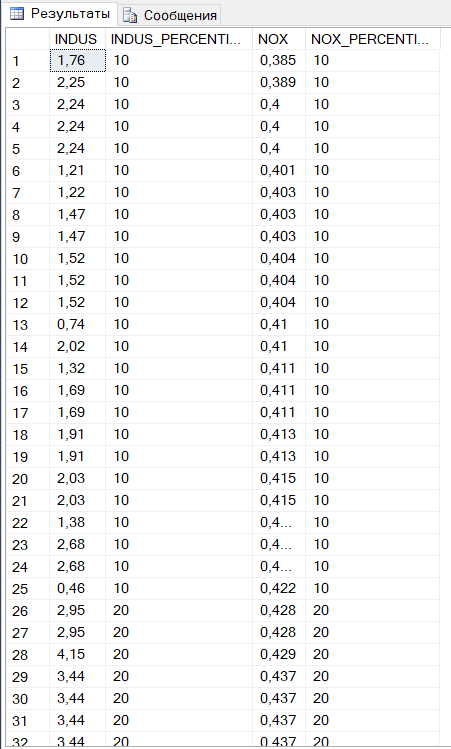
,NOX

,NTILE(10) OVER (ORDER BY NOX) \* 10 AS nox\_percentile

FROM boston

) AS m

WHERE indus\_percentile = nox\_percentile



Прослеживается прямая корреляционная связь между параметрами, т.е при увеличении доли промышленной застройки растет концентрация азота