**Data Cleaning**

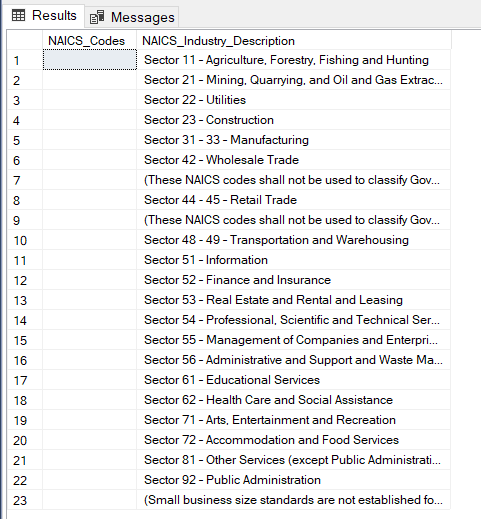
/\*\* Script to remove subsector industry codes, keeping only major sector codes \*\*/

SELECT [NAICS\_Codes]

,[NAICS\_Industry\_Description]

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''



/\*\*\*\*\*\* Script to delete the NAICs code column \*\*\*\*\*\*/

SELECT

[NAICS\_Industry\_Description] AS Industry\_Sector

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''



/\*\* Script to extract industry codes using SQL substring function and add them to a new column \*\*/

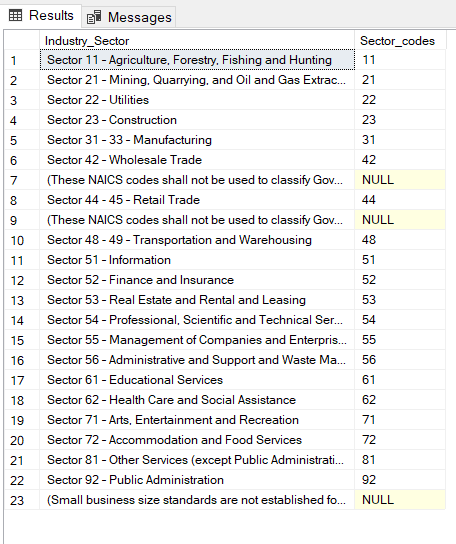
SELECT

[NAICS\_Industry\_Description] AS Industry\_Sector,

CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN SUBSTRING(NAICS\_Industry\_Description, 8, 2) END AS Sector\_codes

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''



/\*\*\*\*\*\* Subquery to drop all NUL values \*\*\*\*\*\*/

SELECT \*

FROM (

SELECT

[NAICS\_Industry\_Description] AS Industry\_Sector,

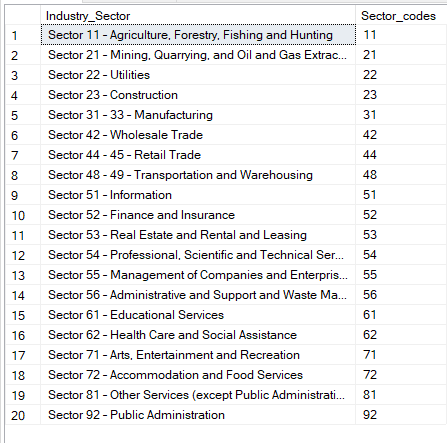
CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN SUBSTRING(NAICS\_Industry\_Description, 8, 2) END AS Sector\_codes

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''

) MAIN

WHERE Sector\_Codes != ''



/\*\*\*\*\*\* Script to extract the Industry Sector Name\*\*\*\*\*\*/

SELECT \*

FROM (

SELECT

[NAICS\_Industry\_Description] AS Industry\_Sector,

CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN SUBSTRING(NAICS\_Industry\_Description, 8, 2) END AS Sector\_codes,

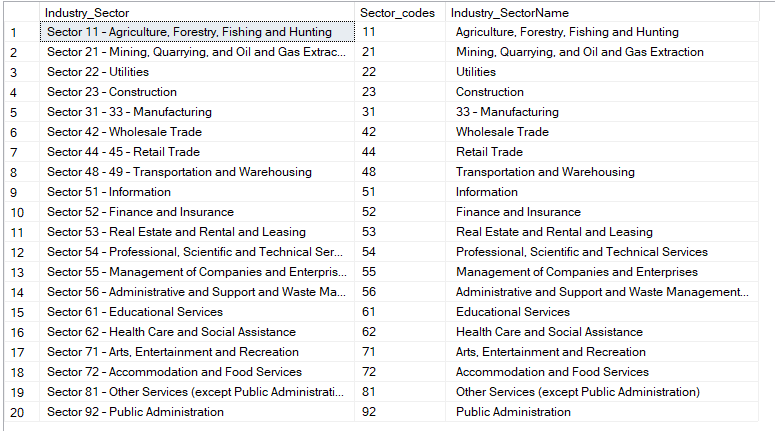
CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN ltrim(SUBSTRING(NAICS\_Industry\_Description, CHARINDEX('–', NAICS\_Industry\_Description) + 1, LEN(NAICS\_Industry\_Description))) END AS Industry\_SectorName

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''

) MAIN

WHERE Sector\_Codes != ''



/\*\* Script to save INTO a new table \*\*/

SELECT \*

INTO SBA\_Sector\_Codes\_Descriptions2

FROM (

SELECT

[NAICS\_Industry\_Description] AS Industry\_Sector,

CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN SUBSTRING(NAICS\_Industry\_Description, 8, 2) END AS Sector\_codes,

CASE WHEN NAICS\_Industry\_Description LIKE '%–%' THEN ltrim(SUBSTRING(NAICS\_Industry\_Description, CHARINDEX('–', NAICS\_Industry\_Description) + 1, LEN(NAICS\_Industry\_Description))) END AS Industry\_SectorName

FROM [SBA].[dbo].[sba\_industry\_standards]

WHERE NAICS\_Codes = ''

/\*\*\*\*\*\* Script to edit and update the table using Insert INTO \*\*\*\*\*\*/

Select

Industry\_Sector,

Sector\_codes,

Industry\_SectorName

from SBA\_Sector\_Codes\_Descriptions

insert into SBA\_Sector\_Codes\_Descriptions

values

('Sector 31 - 33 - Manufacturing', 32, 'Manufacturing'),

('Sector 31 - 33 - Manufacturing', 33, 'Manufacturing'),

('Sector 44 - 45 - Retail Trade', 44, 'Retail Trade'),

('Sector 48 - 49 - Transportation and Warehousing, 48, 'Transportation and Warehousing'),

('Sector 48 - 49 - Transportation and Warehousing, 49, 'Transportation and Warehousing')

Select

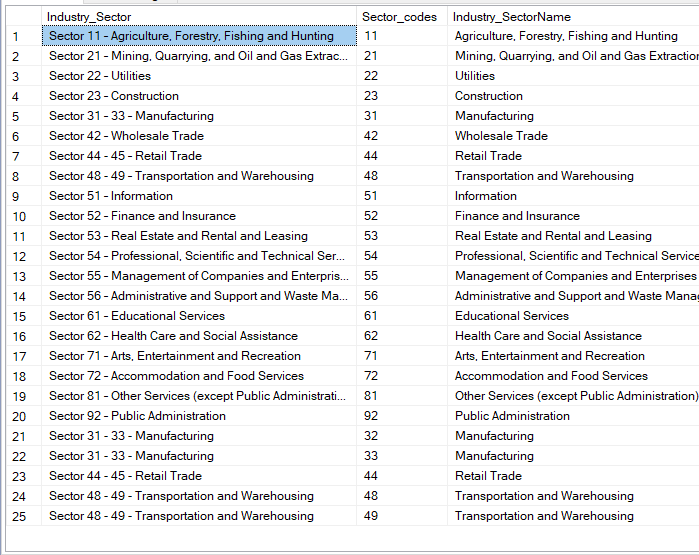
Industry\_Sector,

Sector\_codes,

update SBA\_Sector\_Codes\_Descriptions2

set Industry\_SectorName = 'Manufacturing'

where Sector\_codes = 31



**Exploration and Analysis**

---What is the summary of all the approved PPP loans---

/\*\*\*\*\*\* What are the KPI values for the approved PPP loans? \*\*\*\*\*\*/

SELECT

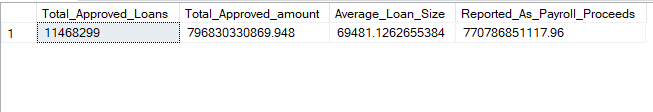
COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

FROM [SBA].[dbo].[SBA\_Public\_Data]



/\*\*\*\*\*\* What percentage of PPP loans were forgiven? \*\*\*\*\*\*/

SELECT

COUNT(LoanNumber) AS Total\_Approved\_Loans,

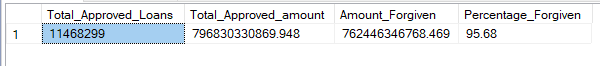
SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

SUM(ForgivenessAmount) AS Amount\_Forgiven,

ROUND(SUM(ForgivenessAmount)/SUM(InitialApprovalAmount) \* 100, 2)As Percentage\_Forgiven

FROM

[SBA].[dbo].[SBA\_Public\_Data]



/\*\*\*\*\*\* KPI’s Grouped by Year \*\*\*\*\*\*/

SELECT

YEAR(DateApproved),

COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

FROM

[SBA].[dbo].[SBA\_Public\_Data]

WHERE

YEAR(DateApproved) = 2020

GROUP BY

YEAR(DateApproved)

UNION

SELECT

YEAR(DateApproved),

COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

FROM

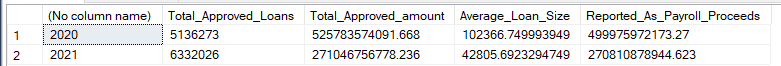
[SBA].[dbo].[SBA\_Public\_Data]

WHERE

YEAR(DateApproved) = 2021

GROUP BY

YEAR(DateApproved)



/\*\* Which are the top 15 lending institutions that provided the majority of PPP loans?

SELECT TOP 15

OriginatingLender,

COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

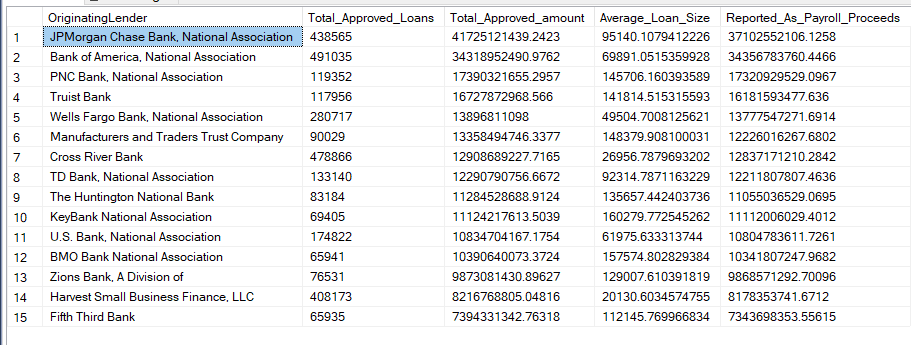
FROM

[SBA].[dbo].[SBA\_Public\_Data]

GROUP BY

OriginatingLender

ORDER BY 3 DESC



/\*\*\*\*\*\* Which are the top 15 Sectors that received the majority of PPP loans? \*\*\*\*\*\*/

SELECT TOP 15

sc.Industry\_SectorName,

COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

FROM

[SBA].[dbo].[SBA\_Public\_Data] pd

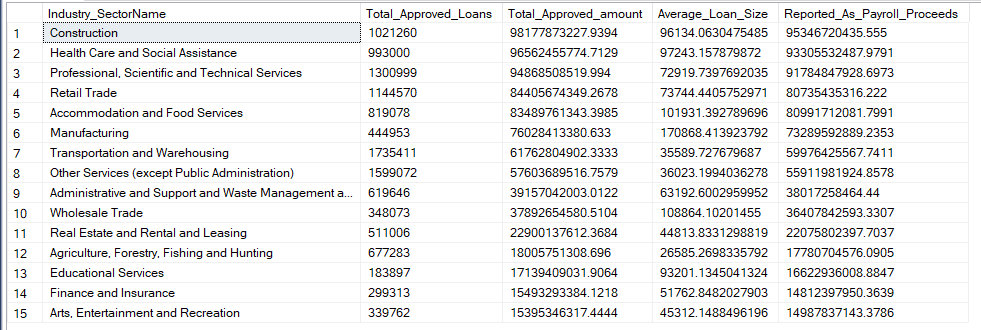
join SBA\_Sector\_Codes\_Descriptions2 sc

ON left(pd.NAICSCode, 2) = sc.Sector\_codes

GROUP BY

sc.Industry\_SectorName

ORDER BY 3 DESC



/\*\*What is the % of approved loans by the top 15 Industry Sectors, using CTE’s \*\*/

; with CTE AS

(

SELECT TOP 15

sc.Industry\_SectorName,

COUNT(LoanNumber) AS Total\_Approved\_Loans,

SUM(InitialApprovalAmount) AS Total\_Approved\_amount,

AVG(InitialApprovalAmount) AS Average\_Loan\_Size,

SUM(PAYROLL\_PROCEED) AS Reported\_As\_Payroll\_Proceeds

FROM

[SBA].[dbo].[SBA\_Public\_Data] pd

join SBA\_Sector\_Codes\_Descriptions2 sc

ON left(pd.NAICSCode, 2) = sc.Sector\_codes

GROUP BY

sc.Industry\_SectorName

--ORDER BY 3 DESC

)

SELECT Industry\_SectorName, Total\_Approved\_Loans, Total\_Approved\_amount,

ROUND(Total\_Approved\_amount/SUM(Total\_Approved\_amount) OVER() \* 100, 2) AS Percentage\_Of\_Total\_Approved\_Amount

from CTE

ORDER BY Total\_Approved\_amount desc

