

```

--Check for any (Duplicates)
SELECT LoanNumber , COUNT(*)
FROM public_all
GROUP BY LoanNumber
HAVING COUNT(*) > 1 ;
-- There are Duplicates

--Removing Duplicates

WITH CTE AS (
    SELECT LoanNumber,
           ROW_NUMBER() OVER (PARTITION BY LoanNumber ORDER BY LoanNumber ASC ) AS RowNum
    FROM public_all
)
DELETE FROM CTE
WHERE RowNum > 1;

-- 1,800,000 Rows Deleted

--CHECK IF THERE IS NULL VALUES IN COLUMN (LoanNumber)

SELECT COUNT(*) FROM public_all
WHERE LoanNumber IS NULL
--35,994

DELETE FROM PUBLIC_ALL
WHERE LoanNumber IS NULL
--35,994 rows affected

```

```

1164 --This query to check all duplicates in all states in table (Borrowers)
1165 SELECT
1166     BorrowerName ,
1167     UPPER(BORROWERADDRESS) AS ADD_CAPITAL ,
1168     UPPER (BorrowerCity) AS CITY_CAPITAL ,
1169     LEFT(BORROWERZIP , 5) AS ZIP_CORRECT ,
1170     JobsReported ,
1171     BorrowerState ,
1172     RuralUrbanIndicator ,
1173     HubzoneIndicator ,
1174     Race ,
1175     Ethnicity ,
1176     Gender ,
1177     Veteran ,
1178     count (*)
1179     FROM Borrowers
1180 group by BorrowerName ,
1181     UPPER(BORROWERADDRESS) ,
1182     UPPER (BorrowerCity) ,
1183     LEFT(BORROWERZIP , 5) ,
1184     JobsReported ,
1185     BorrowerState ,
1186     RuralUrbanIndicator ,
1187     HubzoneIndicator ,
1188     Race ,
1189     Ethnicity ,
1190     Gender ,
1191     Veteran
1192 having count (*) > 1
1193 ORDER BY BorrowerName
1194 --745,676 ROWS

```

```

1196 --Remove duplicates
1197 WITH cte AS (
1198     SELECT
1199     BorrowerName ,
1200     UPPER(BORROWERADDRESS) AS ADD_CAPITAL ,
1201     UPPER (BorrowerCity) AS CITY_CAPITAL ,
1202     LEFT(BORROWERZIP , 5) AS ZIP_CORRECT ,
1203     JobsReported ,
1204     BorrowerState ,
1205     RuralUrbanIndicator ,
1206     HubzoneIndicator ,
1207     Race ,
1208     Ethnicity ,
1209     Gender ,
1210     Veteran ,
1211     ROW_NUMBER () OVER (PARTITION BY UPPER(BORROWERADDRESS) ,
1212     UPPER (BorrowerCity) ,
1213     LEFT(BORROWERZIP , 5) ,
1214     JobsReported ,
1215     BorrowerState ,
1216     RuralUrbanIndicator ,
1217     HubzoneIndicator ,
1218     Race ,
1219     Ethnicity ,
1220     Gender ,
1221     Veteran ORDER BY (SELECT(0))) AS ROW_NUM
1222     FROM Borrowers )
1223
1224     DELETE FROM cte
1225     WHERE ROW_NUM > 1
1226 --1,336,032 ROWS AFFECTED

```

```

1418 -- Calculate the mode
1419 SELECT
1420     top (5) NAICSCode,
1421     COUNT(*) as frequency
1422 FROM
1423     public_all
1424 GROUP BY
1425     NAICSCode
1426 ORDER BY
1427     frequency DESC
1428 -- NAICSCode = 812112 , frequency = 458,411
1429 -- NAICSCode = 722511 , frequency = 326,447
1430 -- NAICSCode = 531210 , frequency = 251,837
1431 -- NAICSCode = 485310 , frequency = 250,850
1432 -- NAICSCode = 812990 , frequency = 231,016
1433
1434 select
1435     *
1436 from
1437     [Size Standards_Effective]
1438 where
1439     NAICS_Codes = '812112'
1440 or NAICS_Codes = '722511'
1441 or NAICS_Codes = '531210'
1442 or NAICS_Codes = '485310'
1443 or NAICS_Codes = '812990'
1444
1445 -- NAICSCode = 812112 , Beauty Salons
1446 -- NAICSCode = 722511 , Full-Service Restaurants
1447 -- NAICSCode = 531210 , Offices of Real Estate Agents and Brokers10
1448 -- NAICSCode = 485310 , Taxi and Ridesharing Services
1449 -- NAICSCode = 812990 , All Other Personal Services

```

```

-----
1591  --lets move to column (BusinessType)
1592
1593  SELECT DISTINCT BusinessType
1594  FROM public_all
1595  -- WE HAVE 25 TYPE OF BUSSINES (PROFIT / NON PROFIT / NULL VALUES )
1596
1597  --AFTER external SEARCH THESE ARE THE NON PROFIT AND (THE OTHER will be PROFIT)
1598
1599  -- 501(c) - Non Profit except 3,4,6,
1600  -- 501(c)19 - Non Profit Veterans
1601  -- 501(c)3 - Non Profit
1602  -- 501(c)6 - Non Profit Membership
1603  -- Cooperative
1604  -- Non-Profit Childcare Center
1605  -- Non-Profit Organization

```

```

1679 SELECT
1680     SUM (InitialApprovalAmount)
1681 FROM
1682     public_all
1683 WHERE
1684     BusinessType = '501(c) - Non Profit except 3,4,6,'
1685     OR BusinessType = '501(c)19 - Non Profit Veterans'
1686     OR BusinessType = '501(c)3 - Non Profit'
1687     OR BusinessType = '501(c)6 - Non Profit Membership'
1688     OR BusinessType = 'Cooperative'
1689     OR BusinessType = 'Non-Profit Childcare Center'
1690     OR BusinessType = 'Non-Profit Organization'
1691
1692 --AMOUNT BORROWED BY (NON PROFIT) = 53,012,523,122 $
1693
1694 SELECT
1695     SUM (InitialApprovalAmount)
1696 FROM
1697     public_all
1698 WHERE
1699     BusinessType <> '501(c) - Non Profit except 3,4,6'
1700     OR BusinessType <> '501(c)19 - Non Profit Veterans'
1701     OR BusinessType <> '501(c)3 - Non Profit'
1702     OR BusinessType <> '501(c)6 - Non Profit Membership'
1703     OR BusinessType <> 'Cooperative'
1704     OR BusinessType <> 'Non-Profit Childcare Center'
1705     OR BusinessType <> 'Non-Profit Organization'
1706 --AMOUNT BORROWED BY (PROFIT) = 790,389,199,292 $
1707
1708 --CONCLUSION : THE (PROFIT) ORGANIZATIONS BORROWED (MUCH MORE) MONEY THAN (NON PROFIT)
1709 --THE (InitialApprovalAmount) START FROM (887 $ : 10M $) WHEN COLUMN (BusinessType) IS NULL
1710
1711 --SO WE WILL REPLACE NULL VALUES IN COLUMN (BusinessType) DEPENDS ON COLUMN (InitialApprovalAmount)
1712 --IF THE (InitialApprovalAmount) ABOVE 4M $ IT WILL BE ONE TYPE OF PROFIT ORGANIZATION
1713 --IF THE (InitialApprovalAmount) BELOW 4M $ IT WILL BE ONE TYPE OF NON PROFIT ORGANIZATION
1714

```

```

1826 --we want to handle null values in column (NonProfit) based on column (BusinessType)
1827 --and also replace 'Y' with Non Profit
1828
1829 update public_all
1830 set NonProfit =
1831     CASE
1832         WHEN BusinessType = '501(c) - Non Profit except 3,4,6,' THEN 'Non Profit'
1833         WHEN BusinessType = '501(c)19 - Non Profit Veterans' THEN 'Non Profit'
1834         WHEN BusinessType = '501(c)6 - Non Profit Membership' THEN 'Non Profit'
1835         WHEN BusinessType = 'Non-Profit Childcare Center' THEN 'Non Profit'
1836         WHEN BusinessType = 'Cooperative' THEN 'Non Profit'
1837         WHEN BusinessType = '501(c)3 - Non Profit' THEN 'Non Profit'
1838         WHEN BusinessType = 'Non-Profit Organization' THEN 'Non Profit'
1839         ELSE 'Profit'
1840     END;
1841
1842 --(11,335,950 rows affected)

```

```

1950
1951 --Analyze Part
1952
1953 --Calculate top & bottom (10) Total Loans & Total Amount for each industry
1954 SELECT
1955     top (10) p.NAICSCode,
1956     COUNT(p.loannumber) as tot_loans,
1957     round(
1958         SUM(p.initialapprovalamount),
1959         2
1960     ) as tot_amount,
1961     s.NAICS_Industry_Description
1962 FROM
1963     public_all as p
1964     join [Size Standards_Effective] as s on p.NAICSCode = s.NAICS_Codes
1965 group by
1966     p.NAICSCode,
1967     s.NAICS_Industry_Description
1968 order by
1969     tot_amount desc

```



```
2008  --sum of amount given by top & bottom (10) servicing lender(Bank) & given to which industry
2009
2010  select
2011      top (10) p.ServicingLenderLocationID,
2012      s.ServicingLenderCity,
2013      s.ServicingLenderName,
2014      z.NAICS_Industry_Description,
2015      sum(p.InitialApprovalAmount) as TOT_amount
2016  from
2017      public_all as p
2018      join ServicingLender as s on p.ServicingLenderLocationID = s.ServicingLenderLocationID
2019      join [Size Standards_Effective] as z on p.NAICSCode = z.NAICS_Codes
2020  group by
2021      p.ServicingLenderLocationID,
2022      s.ServicingLenderCity,
2023      s.ServicingLenderName,
2024      z.NAICS_Industry_Description
2025  order by
2026      TOT_amount desc
```

```
2103 select
2104     SUM(InitialApprovalAmount) as TOT_amount,
2105     month(dateapproved) as Month,
2106     year(dateapproved) as Year
2107 from
2108     public_all
2109 group by
2110     month(dateapproved),
2111     year(dateapproved)
2112 having
2113     year(dateapproved) is not null
2114 order by
2115     year(dateapproved),
2116     month(dateapproved)
```