**1) We imported large amount of csv file into sql quickly using load data infile**

**2)check the formats before importing into sql to avoid data loss**

**a) data default format in sql is yyyy-mm-dd**

**b)normal number formats (don’t use percentage,currency ,accounting formats because they make sql read as a text or string rather number)**

**c) don’t use , (thousand separator for numbers)**

**3)check for any duplicates that exists in excel . when check we don’t see any duplicates . Select entire file > data > remove duplicates**

**4) check for missing values in each column by using formula ( =COUNTIF(A3:A9242, "") / COUNTA(A3:A9242))**

**5)Eliminated the columns that has more than 50% of data missing as it may not help in delivering accurate insights**

**Tasks done on SQL**

1. **Check the total rows and columns are loaded using the command**
2. **Then check for null values ,missing or blanks**
3. **Drop the columns irrelevant and has more than 50 % missing values (Tags,Lead Quality,Asymmetric activity index and profile index) can drop these**
4. **Using SET SQL\_SAFE\_UPDATES = 0; we disable safe mode and update the table**
5. **Select each column and check if we need to go with mode,median or mean and update it with those values**
6. **Check how many blank counts and null counts are there**
7. **Similarly we treated for all columns as per data with column names**

**#from excel we can see lead\_source,Totalvisits,page views per visit, last activity,country,specialisation, how did you hear about x education, what is your current occupation,#what matters to you most choosing a course,tags, lead\_quality,lead\_profile,city,asymmetric activity index,asymmetric profile index,asym activity score, asym profile score have blank or null values or spaces**

**8)** **to calculate mean we run**

**# SELECT AVG(column\_name) FROM cust\_leads WHERE column\_name IS NOT NULL;then use update statement**

**9)dump the entire cleaned data into outfile from SQL for POWER BI ANALYSIS**

**Tasks done on Power BI**

**1)replaced with India under power query for Mumbai, Thane and outskirts, other cities of Maharashtra region as part of data cleaning**

**2)Under table tools we created new measures rather columns for total leads,converted leads, conversion ratio.**

**Corrected Country =**

**SWITCH(**

**TRUE(),**

**SEARCH("Mumbai", 'Cleaned data set'[City], 1, 0) > 0, "India",**

**SEARCH("Thane", 'Cleaned data set'[City], 1, 0) > 0, "India",**

**SEARCH("Other Cities of Maharashtra", 'Cleaned data set'[City], 1, 0) > 0, "India",**

**'Cleaned data set'[Country]**

**)**

**3) developed various visuals based on lead conversion , conversion ratio etc…**

**4)Prepared final dashboard**

**Final Analysis Insights from Dashboard**

**1)We have total 9240 Leads,3561 converted leads , avg page view per visit is 2.37,avg time spent on website is 487.70 and conversion rate is 38.54%**

**2)We have blog,live chat,NC\_EDM,pay per click ads,Press release,social media,testone,welearn,welearn blog home having least number of total leads which are 1 or 2**

**3) Top 4 Lead sources are Google, Direct Traffic, Olark Chat, Organic search**

**4) Leads by city top 3 are Mumbai,Unknown,Thane & Outskirts**

**5)Leads by country top 3 are India,Unknown and united Arab Emirates**

**6) Through Lead Add form we have 92.48% conversion rate with 718 total leads highest through reference as lead source and last notable activity is SMS sent and top 3 specialisations are unknown,marketing and human resource management**

**7) then we have Landing Page Submission with 4886 total leads,1768 converted leads 36.19% as conversion ratio where last notable activity is sms sent,email opened and specialisation in finance and human resource management**

**8)Quick Add Form and Lead Import are the least Lead origins with total leads as 1 and 55 with conversion rates as 100% and 23.64% and unemployment and unknown are major occupations,and most of lead sources are facebook**

**9) we have 3561 out of 9240 as converted leads with avg page views per visit is 2.33 and avg time spent on website is 738.55 where top 2 occupations are unemployment and working professionals with last notable activity is SMS sent and email opened, highest lead source is google and direct traffic and top 3 specialisations are unknown,finance and marketing management and most leads are from India followed by unknows and united Arab Emirates.**

|  |  |
| --- | --- |
|  |  |