

Big Data Analytics Final Project

Project can be completed using both spark and hive.

Project contains covid19 data for PHU locations and cases in Canada.

Project folder has 2 files 1) phu_locations.json 2) cases_by_status_and_phu.csv

Try to find following questions.

- Task 1 to 8 should be done by using spark (dataframe or dataset or sparksql or rdd).
- Task 9 to 15 should be done by using both hive and spark.

Questions :

Task 1 : list all PHU locations which are allow online appointments. List should not contain columns in French. All French columns have name ending with fr.

Task 2 : List all the PHU locations allow children under 2 in Brampton location.

Task 3: List all the PHU locations in Toronto has free parking with drive through.

Task 4 : Find out number of PHUs are closed in each city on Friday

Task 5 : find out all the PHU locations which are open during 20:00 and 21:00 on Monday in Toronto city.

Task 6 : List the hours of operation of all the PHU in Toronto. Use below format.

PHU	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
York Region PHU	8:00 to 5:00	8:00 to 5:00	8:00 to 5:00	8:00 to 5:00	8:00 to 5:00	8:00 to 5:00	8:00 to 5:00

Task 7 : Count all the PHU are temporary closed in each city.

Task 8 : phu_locations.json has phone field. Create two more columns. First field should contain only area code and 2nd field should contain extension number.

phone	area_code	extension
416-586-4800 ext. 4123	416	4123

Task 9 : Find out highest number of resolved cases of covid-19 in each PHU between April 2020 to Sep 2020

Task 10 : which PHU has more resolved cases of covid-19 than active cases in year 2020.

Task 11 : List only PHU which has least death due to covid-19 in each month.

Task 12 : Find out total active cases remaining in each PHU

Task 13 : Find out on which date highest and lowest death reported in “NIAGARA REGION” PHU.

Task 14 : Find out total resolved cases of covid-19 in “NORTH BAY PARRY SOUND DISTRICT” in month May 2020 and Oct 2020.

month	resolved_cases
May2020	15285
Oct2020	29333

Task 15 : Find out percentage of active cases in each PHU in year 2020. Result should be link below.

PHU_NAME	total_percentage
HALTON REGION	25.34%
HURON PERTH	12.00%
LAMBTON COUNTY	6.34%
NORTHWESTERN	2.33%
TIMISKAMING	7.59%
SUDBURY AND DISTRICT	19.02%