Kaggle Dataset
==============
https://www.kaggle.com/openaddresses/openaddresses-europe

create a directory in HDFS to collect the input data:
==
hadoop fs -mkdir /user/hive/online_sales

Import File from S3
==
aws s3 ls e-commerce-events-ml

Check Directory
==
hadoop fs -ls /user/hive

To move datasets from S3 to HDFS:
===================================
hadoop distcp 's3://e-commerce-events-ml/*' '/user/hive/online_sales/'

To Check the file
===================================
hadoop fs -cat /user/hive/online_sales/2019-Nov.csv | head

hadoop fs -cat /user/hive/online_sales/2019-Oct.csv | head

In new Terminal
==
==================

Use Hive command
================================
hive

Create & Use database
================================
create database online_sales;
 use online_sales;

Create base table to read the input data:
==

CREATE EXTERNAL TABLE IF NOT EXISTS sales_input(event_time timestamp ,event_type string ,product_id string ,category_id string ,category_code string ,brand string ,price float ,user_id bigint ,user_session string)

ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
STORED AS TEXTFILE
LOCATION '/user/hive/online_sales/'
tblproperties("skip.header.line.count"="1");

To enable partitioning and bucketing:
====================================

set hive.exec.dynamic.partition.mode=nonstrict;
set hive.exec.dynamic.partition=true;
set hive.enforce.bucketing=true;

CREATE TABLE IF NOT EXISTS sales_bucket(event_time timestamp, product_id
string, category_id string, category_code string, brand string, price float, user_id
bigint, user_session string)
PARTITIONED BY (event_type string) CLUSTERED BY (price) into 10 buckets
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
STORED AS TEXTFILE;

To load the optimized hive table:
===============================

insert into table sales_bucket partition(event_type) select event_time, event_type,
product_id, category_id, category_code, brand, price, user_id, user_session from
sales_input;

Performance comparison:
=======================
select * from sales_input where event_type = "cart" and price > 300;

select * from sales_bucket where event_type = "cart" and price > 300;

Required Questions
==
================

Find the total revenue generated due to purchases made in October.
==
====

select sum(price) from sales_bucket where event_type = 'purchase' and month(event_time)
= 10;

Output:
1211538.430000433

Write a query to yield the total sum of purchases per month in a single output.
==
====

```sql
select sum(price) from sales_bucket where event_type = 'purchase' group by
month(event_time);
```

Output:
1211538.430000433
1531016.9000001205

Write a query to find the change in revenue generated due to purchases from
October to November.
===
==============

```sql
select Oct, Nov, Nov - Oct Difference
from
(
SELECT sum(case when date_format(event_time,'MM')=10 then price else 0
end) AS Oct,
     sum(case when date_format(event_time,'MM')=11 then price else 0 end)
AS Nov
 FROM sales_bucket WHERE date_format(event_time,'MM')in (10,11) AND
event_type='purchase'
)s;
```

Output:
1211538.430000433   1531016.9000001205 319478.4699996875

Find distinct categories of products. Categories with null category code can be
ignored.
===
====

```sql
select distinct category_code from sales_bucket;
```

Output:
accessories.cosmetic_bag
stationery.cartrige
accessories.bag
appliances.environment.vacuum
furniture.living_room.chair
sport.diving
appliances.personal.hair_cutter
appliances.environment.air_conditioner
apparel.glove
furniture.bathroom.bath
furniture.living_room.cabinet

Find the total number of products available under each category.
===
====

select category_id, count(product_id) from sales_bucket group by category_id;

Output:

1487580004966466385	16
1487580005050352469	83278
1487580005176181595	127
1487580005369119587	3
1487580005570446188	24
1487580005671109489	300570
1487580005687886706	14
1487580005922767741	640
1487580006056985476	1794
1487580006073762693	7556
1487580006174425994	466
1487580006216369036	3
1487580006585467807	17
1487580007281722301	34854
1487580007432717250	64400
1487580007508214725	36
1487580007592100809	12450
1487580007852147670	42694
1487580007894090712	4957
1487580007910867929	51488
1487580007952810971	24742
1487580008053474272	2629
1487580008070251489	1643
1487580008087028706	778
1487580008221246441	46
1487580008472904691	65
1487580008523236341	11
1487580009227878422	12922
1487580009336930331	5784
1487580009605365797	39093
1487580010955931741	835
1487580011098538083	3368
1487580011224367209	1
1487580011283087468	28420
1487580011408916594	20134
1487580011476025461	1871
1487580011677352062	17310
1487580011702517887	30290
1487580011970953351	4128
1487580012121948301	4731
1487580012147114126	367
1487580012507824279	101
1487580012784648356	1328
1487580012851757223	10
1487580012952420523	1948
1487580013363462335	3
1487580013430571202	5787
1487580013506068678	17853
1487580013749338323	31471
1487580013900333275	2938

1487580013992607968	1137
1487580014042939619	213
1516331853567492962	5632
1554383493545328915	1600
1605161575889502297	15760
1645114480121610699	6353
1783999064170299632	109
1783999072332415142	34726
1783999073758478650	6663
1889472915104072007	1155
1891434214553813877	25792
1897124495995438074	3570
1933364344720982871	2808
1998040850688377703	1017
2018395024110125980	6689
2020400296164851991	614
2022516588854378951	723
2028340285225828822	314
2071303198680810125	802
2084144451428549153	85721
2094448780651791052	825
2095736144888071137	2092
2106514244437541443	1422
2114584564549550293	16995
2121383893343929118	870
2151191059827262021	332
2151191075757228942	1088
1487580004857414477	47064
1487580005008409427	33512
1487580005092295511	321824
1487580005411062629	55670
1487580005486560104	2140
1487580005654332272	3
1487580006157648777	592
1487580006317032337	283387
1487580006644188066	20307
1487580006744851367	5
1487580007021675441	2926
1487580007046841266	3625
1487580007113950133	1343
1487580007189447609	10664
1487580007306888126	28974
1487580007399162817	14263
1487580007701152718	853
1487580007717929935	40841
1487580007759872977	3538
1487580007835370453	6844
1487580008011531230	3547
1487580008128971748	100
1487580008204469224	2
1487580008263189483	72932
1487580008313521133	29562
1487580008422573041	1486
1487580008590345208	124

1487580008984608779	11702
1487580009076883471	7344
1487580009118826513	5005
1487580009471148064	60473
1487580009521479714	12066
1487580009546645539	3052
1487580009622143014	5292
1487580009739583530	329
1487580009764749355	9
1487580009798303788	128
1487580009857024046	1
1487580009974464562	643
1487580010075127862	61
1487580010100293687	85154
1487580010125459512	29435
1487580010150625337	184
1487580010217734204	3817
1487580010377117763	2936
1487580010645553231	39081
1487580010737827924	2683
1487580010754605141	7670
1487580010897211483	1
1487580010989486174	13
1487580011308253293	11029
1487580011333419118	750
1487580011601854587	736
1487580011903844485	1792
1487580011996119176	5712
1487580012071616651	162
1487580012205834384	2606
1487580012373606546	8941
1487580012432326804	4631
1487580012482658454	39
1487580012524601496	7133
1487580012549767321	2
1487580012642042013	9585
1487580012927254698	29008
1487580012994363565	18548
1487580013069861041	39309
1487580013145358517	9759
1487580013254410426	8325
1487580013564788937	112
1487580013657063630	3270
1487580013690618064	6601
1487580013841613016	148361
1487580014009385185	10968
1495705810754339379	6
1511892746070131099	27234
1558526315760452545	2379
1559261858748170632	1495
1584505206060614455	707
1588205336064425989	1996
1628099626433249931	187
1648815651034235876	2167

1648815843896722076	7
1660394382047576308	1555
1720400165430363096	4955
1761186209054327497	5797
1783999068909863670	35339
1783999072365969578	2448
1791442703716712702	22
1791442849384891169	894
1791442895991997390	1839
1797122099068797194	1188
1804383582576181309	2522
1805953852441101266	1965
1871018959927509750	6157
1891434351850160381	9257
1911999801642189621	2719
1921723506584715388	1248
1926797403503985079	1322
1933472286753424063	5288
1962525126503170858	492
1977575787259232875	8293
1998040852064109417	3904
2013754353822728372	2821
2022622168218599898	308
2055368408169447599	1668
2060156961931919712	112
2068966806634103136	538
2130081478220972046	1464
2151191059751764547	2140
2151191071378375538	36371
2154396123597373922	503
2187686850687140020	673
2187790129827939246	86
2195085258272014535	3880
1487580004983243602	556
1487580005067129686	14721
1487580005268456287	194193
1487580005293622112	582
1487580005318787937	211
1487580005427839846	102994
1487580005461394279	61348
1487580005595612013	322269
1487580005629166447	2030
1487580005796938615	1
1487580005855658874	7011
1487580005998265217	242
1487580006098928518	1636
1487580006132482952	15822
1487580006451250072	5358
1487580006509970331	17967
1487580006526747548	5155
1487580006610633632	46
1487580006820348840	1223
1487580006870680490	2251
1487580007004898224	1015

1487580007214613434	4938
1487580007575323592	28439
1487580007776650194	12346
1487580007877313495	3210
1487580008162526182	9084
1487580008288355308	27172
1487580008447738866	21129
1487580008657454075	32679
1487580008758116352	8597
1487580008816836611	4527
1487580009051717646	62698
1487580009143992338	39775
1487580009202712597	12482
1487580009387261981	44756
1487580009571811364	5081
1487580010268065854	4144
1487580010293231679	3077
1487580010335174721	3534
1487580010561667147	23171
1487580010628776014	30796
1487580010662330448	1636
1487580010721050707	3631
1487580010821714008	17757
1487580011031429216	546
1487580011425693811	18511
1487580011752849537	904
1487580012021285001	14008
1487580012591710363	5451
1487580012717539489	4930
1487580012734316706	523
1487580012759482531	249
1487580012809814181	942
1487580012902088873	28499
1487580012969197740	16859
1487580013053083824	26722
1487580013170524342	9748
1487580013279576251	17752
1487580013296353468	4338
1487580013338296510	2948
1487580013413793985	12246
1487580013732561106	8626
1487580013782892757	11
1487580013858390233	27189
1487580013950664926	122970
1547480590851244887	674
1597769965795017114	4345
1602943681873052386	245975
1638456119066100510	64214
1657722039387029886	380
1752742610406999036	529
1752742615205281895	5245
1759024698982662957	235
1759024699007828783	2657
1783999063574708423	3673

1783999064103190764	12150
1783999064136745198	24139
1783999067156644376	38699
1783999073792033086	3136
1783999076535108182	1174
1783999076551885399	6872
1805953965678920077	34
1814592470537732872	2715
1819693959081886239	7269
1842735760499802745	10491
1842735768846467167	988
1858245586344477369	3400
1897124469017673788	3608
1897124489259385629	1818
1911999801088541491	3325
1911999948073730325	684
1924049106385240809	14340
1937169073007756269	8852
1962525462299148572	344
1977786601392047073	1690
1982860244379763042	7156
1987724780047958178	248
2018287324474901238	1401
2027602240612598357	388
2055161088059638328	14940
2140803113261466607	18058
2145935122136826354	305
2151191070908613477	7448
2151191070984110951	9168
2151191071118328683	13351
2164688961165852944	229
2195085255176618020	4009
1487580004832248652	25536
1487580004882580302	25569
1487580004916134735	103859
1487580005025186644	1596
1487580005880824699	2117
1487580005897601916	13385
1487580005939544958	13
1487580006015042434	2009
1487580006199591819	6488
1487580006241534861	13760
1487580006300255120	124460
1487580006350586771	59761
1487580006409307030	10368
1487580006484804506	1213
1487580006551913373	25317
1487580006627410849	2163
1487580006711296933	1233
1487580006845514665	792
1487580007139115958	1480
1487580007172670392	4932
1487580007365608384	6331
1487580007483048900	18580

1487580007524991942	2858
1487580007634043851	58516
1487580007659209676	16781
1487580007675986893	497756
1487580007936033754	34352
1487580008187692007	38132
1487580008565179383	1
1487580008607122425	19843
1487580008774893569	9068
1487580008909111303	4951
1487580008925888520	29
1487580009026551821	8979
1487580009177546772	11805
1487580009286598681	151845
1487580009311764506	59213
1487580009362096156	15979
1487580009445982239	102738
1487580010024796212	2
1487580010049962037	4
1487580010242900029	4540
1487580010318397504	8770
1487580010360340546	154
1487580010695884882	11681
1487580010796548183	27525
1487580010872045658	19868
1487580010922377308	2693
1487580011006263391	9
1487580011115315300	8
1487580011182424167	1795
1487580011241144426	1479
1487580011383750769	28
1487580011652186237	34628
1487580011828347011	47
1487580011945787526	849
1487580012096782476	49323
1487580012231000209	1
1487580012658819230	2
1487580012683985055	7262
1487580013027917999	12920
1487580013086638258	6214
1487580013128581300	3616
1487580013212467384	7442
1487580013229244601	9931
1487580013321519293	3030
1487580013489291461	15722
1487580013581566154	47720
1487580013615120588	4811
1487580013824835799	5351
1487580013917110492	2963
1495705810704007729	3880
1525995662934540829	19622
1526733091857498510	7408
1532652067498229876	3704
1543705961326182546	3434

1597770225539875791	13740
1604427145138930210	107
1715102762414375164	2233
1752742606699234159	2355
1783999064036081896	208
1783999068867920626	16291
1783999072407912622	1480
1801635223591453292	2349
1842735758805303837	11080
1842735764450837316	68
1842735764509557576	3128
1913479463425802701	213
1921723491720102387	5733
1924049287554007223	821
1936327050549789532	297
1944358258223350494	1220
1962525118928257818	206
1977575775473238831	323
2029082628195353599	7775
2029731308699124089	4466
2069804417665728971	9383
2069804424703771380	215
2089259162625114209	7438
2115334439910245200	38697
2134354342373753638	9148
2151191071051219817	37008
2155132423103316327	248
2177933350667289121	5597
2193074740619379535	13439
2193074740686488401	3712
2195085255034011676	23587
2195085255117897760	2085
1487580005134238553	163722
1487580005343953762	2953
1487580005385896804	9169
1487580005511725929	110421
1487580005528503146	16249
1487580005553668971	63219
1487580005713052531	119563
1487580005754995573	145435
1487580006434472855	10778
1487580006895846315	46311
1487580006937789357	19490
1487580006979732399	1777
1487580007072007091	4205
1487580007097172916	1798
1487580007256556476	50017
1487580007457883075	9805
1487580007550157767	10005
1487580007986365405	3229
1487580008028308447	2619
1487580008112194531	9599
1487580008145748965	122989
1487580008246412266	157042

1487580008674231292	85
1487580008699397117	3907
1487580008800059394	25151
1487580008858779653	828
1487580008951054345	157
1487580009009774604	8
1487580009261432856	10290
1487580009496313889	31741
1487580009647308839	1
1487580009672474664	2744
1487580009823469613	685
1487580009890578479	1
1487580010192568379	1310
1487580010427449413	3
1487580010603610189	7
1487580010779770966	8321
1487580010846879833	912
1487580011140481125	1
1487580011157258342	4979
1487580011199201384	6
1487580011501191286	15739
1487580011517968503	2774
1487580011534745720	11974
1487580011559911545	5
1487580011585077370	88000
1487580011627020412	31022
1487580011786403970	1259
1487580011853512836	13464
1487580012172279951	2100
1487580012457492629	1474
1487580012574933146	10450
1487580012616876188	1146
1487580012876923048	1608
1487580013011140782	29205
1487580013388628160	20437
1487580013472514244	7463
1487580013522845895	70148
1487580013539623112	17687
1487580013640286413	18309
1487580013799669974	3316
1487580013933887709	376
1487580014093271270	523
1495705810662064688	10994
1495705810729173554	5433
1542195323827388674	19035
1604427094756950459	2397
1658462125284131265	49543
1715102773747384334	3045
1725504706412807026	1819
1752742617696698537	25
1783999063314661546	39932
1783999067181810204	286
1783999071199952917	19733
1783999071325782053	5830

```
1810470908326838736      225
1897124478404526487      22374
1911999884991397970      394
1924049110428549877      20300
1958278551207674674      4957
1982860263572898112      6507
1998040849203594085      5995
2007399943458784057      18070
2035665444290953519      7792
2069171133327868014      2028
2093602042093240877      3188
2106514244487873093      1472
2134354356349173879      257
2141560642253881670      12861
2166295400451933025      11
2193074740493550411      1749
2193074740552270669      13772
2195085258339123402      25
```

Which brand had the maximum sales in October and November combined?
==
====

select brand, sum(price) as pr from sales_bucket where event_type = 'purchase' group by brand order by pr desc limit 2;

Output:

```
        1094188.3000002217
runail148297.94000001193
```

Which brands increased their sales from October to November?
==
====

select brand, sum(price) from sales_bucket where event_type = 'purchase' and (month(event_time) = 10) < (month(event_time) = 11) group by brand;

Output:

```
        619509.2399999945
airnails     5691.5200000000095
almea        973.8700000000003
ardell843.6500000000003
art-visage   2997.8000000000056
artex 4327.249999999993
aura  177.50999999999996
balbcare     212.3799999999997
barbie       12.39
batiste      874.1699999999998
beautix      12222.949999999979
beauty-free1782.8600000000104
beautyblender     109.40999999999998
```

beauugreen	768.3499999999998
benovy	3259.970000000003
bergamo	144.3
bespecial	70.5
binacil	24.259999999999998
bioaqua	1398.1200000000001
biofollica	257.93999999999994
biore	90.30999999999999
blixz	63.4
bluesky	10565.529999999775
bodyton	1380.6399999999999
bpw.style	14837.440000000077
browxenna	14916.730000000072
candy	799.3799999999997
carmex	243.35999999999999
chi	538.61
cnd	29166.589999999946
coifin	1428.4899999999998
concept	13380.399999999938
consly	153.85000000000002
coocla	133.0
cosima	20.929999999999996
cosmoprofi	14536.990000000053
coxir	185.05
cristalinas	584.9499999999999
cruset	145.28
cutrin	367.62
de.lux	2775.509999999994
deoproce	329.16999999999996
depilflax	2803.779999999999
dermal	257.08
dewal	61.29
dizao	945.5099999999994
domix	12009.169999999984
dr.gloderm	11.07
ecocraft	241.95000000000005
ecolab	1214.3000000000002
egomania	146.04000000000002
elizavecca	204.3
ellips	606.04
elskin	307.6500000000001
emil	4098.82
enas	14.1
enigma	224.85000000000002
enjoy	136.57
entity	719.2599999999996
eos	152.61
estel	24142.66999999991
estelare	471.87
eunyul	234.15000000000026
f.o.x	8577.279999999999
fancy	50.620000000000005
farmavita	1291.9699999999996
farmona	1843.4299999999996

farmstay 1074.0599999999997
fedua 263.81
finish 230.38
fly 27.17
foamie 80.49
freedecor 7671.799999999991
freshbubble 502.3399999999998
frozen 12.18
gehwol 1557.6800000000003
glysolid 91.58999999999999
godefroy 425.12
grace 102.60999999999999
grattol 71472.7100000016
greymy 489.48999999999984
happyfons 1091.5900000000006
haruyama 12352.910000000102
helloganic 3.1
i-laq 366.71999999999986
igrobeauty 645.0700000000004
ingarden 33566.210000000065
inm 351.2099999999999
inoface 70.47
insight 1721.9600000000005
irisk 46946.03999999892
italwax 24799.369999999755
jaguar 1110.65
jas 3657.430000000002
jessnail 33345.230000000425
joico 2015.1
juno 21.08
kaaral 5086.070000000007
kamill 81.49000000000001
kapous 14093.080000000027
kares 59.45
kaypro 3268.699999999999
keen 435.62
kerasys 525.2
keune 375.1
kims 632.0400000000001
kinetics 6945.260000000038
kiss 817.3300000000002
kocostar 594.9300000000001
koelcia 112.75
koelf 507.28999999999985
konad 810.6699999999987
koreatida 46.11
kosmekka 1813.37
labay 41.78
laboratorium 312.52
lador 2471.5300000000007
ladykin 170.57
lakme 602.19
lamixx 672.2600000000001
latinoil 384.59000000000003

lebelage	218.27000000000004
levissime	3085.3100000000027
levrana	3664.100000000002
lianail	16394.240000000096
likato	340.97
limoni	1796.6000000000004
litaline	135.74
lovely	11939.059999999932
lowence	567.75
lsanic	959.3399999999999
mane	260.26
marathon	10273.099999999999
markell	2834.430000000003
marutaka-foot	109.33000000000001
masura	33058.469999999696
matreshka	182.67000000000002
matrix	3726.7400000000002
mavala	446.32000000000005
max	8664.230000000001
meisterwerk	340.01000000000005
metzger	6457.160000000007
milv	5642.009999999955
miskin	293.06999999999994
missha	2150.2799999999997
moyou	10.280000000000001
nagaraku	5327.68
naomi	389.0
naturmed	67.31
nefertiti	366.64
neoleor	51.7
nirvel	234.32999999999998
nitrile	1162.679999999999
nitrimax	1809.6400000000035
oniq	9841.650000000023
orly	931.09
osmo	762.31
ovale	3.1
parachute	307.30999999999995
petitfee	864.7999999999997
philips	6.86
plazan	194.01
pnb	6372.489999999998
polarus	11371.930000000008
pole	5527.239999999981
profepil	118.02000000000001
profhenna	736.8499999999996
protokeratin	456.7900000000001
provoc	1063.8200000000013
rasyan	28.939999999999998
refectocil	3475.5800000000063
riche	202.41000000000003
rocknailstar	1.9
rosi	3841.560000000005
roubloff	4913.770000000014

```
runail   76758.65999999491
s.care       913.0699999999999
sanoto       1209.6799999999998
sawa 45.5
severina    6120.479999999983
shary 1176.4899999999993
shik   4839.720000000007
siberina     337.6499999999999
skinity      12.440000000000001
skinlite     890.4500000000008
skipofit     8.49
smart 5902.140000000007
soleo 212.52999999999966
solomeya   2685.8000000000025
sophin       1515.5200000000013
staleks      11875.61000000004
strong       38671.26999999998
sun    65.9
sunuv        8042.150000000003
supertan   66.50999999999999
swarovski   3043.1600000000053
tannymaxx 171.28
tazol 7.18
tertio 245.79999999999998
thuya 2604.9400000000005
tosowoong 27.3
treaclemoon      181.49
trind   542.9600000000002
uno    51039.749999999956
uskusi       5690.31000000004
veraclara   71.21
vilenta      231.20999999999998
vosev        316.7
weaver       6.48
yoko  11707.879999999994
ypsed        436.32
yu-r   673.7099999999999
zeitun       2009.6299999999994
zinger       6684.86000000003
```

Your company wants to reward the top 10 users of its website with a Golden Customer plan.
Write a query to generate a list of top 10 users who spend the most.
===
====

```
select user_id, sum(price) as pr from sales_bucket where event_type = 'purchase' group by
user_id order by pr desc limit 10;
```

Output:

```
557790271 2715.87
150318419 1645.9699999999998
562167663 1352.85
```

531900924 1329.4499999999998
557850743 1295.48
522130011 1185.3899999999996
561592095 1109.7000000000003
431950134 1097.5899999999995
566576008 1056.3600000000006
521347209 1040.9099999999996