PO	PU° Date 22/03/2027
	Base No. 32
Topic	
	Experiment-10
	- R. A
9	CSV que manifulation
1	import file veine incidents, csv into Databerick
	File System.
2	Import SQLContext
	from pyspark, sge import S& Context
	1
3	load the file and create DataFrame.
\rightarrow	crimeincidentOF = spark. read. csv (" dups: / Filestone/
	tables/orine incidents. CSN "inferSchema = "tome"
	neadre = (toure?)
	Collect
->	vineincident DF. Wellect C
2	Take
-	Vineincident Df. take (2)
₹.	Show
-	crimeincident DF. showC
7	shows first & records in tabular format.
->	Vineincident pF. shav (2)
•	
2	Rinding no. of records.
7	Crimeineidentof, court
LAMAL	Teacher's Signature

OPU	Date
aig	Page No33
g brinting the first now of a data > vineincidentor first	
	r Million Committee
10 Schema of dataframe	November 1
10 Schema of dataframe.	
11 Rescribe dataframe	The state of the s
11 <u>Pescuire dataframe</u> > vinneincident DF, describe (), show()
Select function. > new of = crimeincident of, select ('In "lategory', ' new of, show)	1 7 - 1 - 5 - 5 - 5 - 5 - 5 - 5 - 5 - 5 - 5
= new DF = vime incident DF, select (II	reidht Niem?
"Cartegory",	Describt's
new DF, enow	
13 Selectina all collemns.	·
13 Selecting all columns. news = vineindidents, select 1th	27) 4 1 11 12 11 WE
newor, show()	7 - 1 - 1 - 1
	* 1 ·
14 (ovelation between 2 fields.	
Voulation between 2 fields.	2
7	•
E Remone Ouplicate records.	
> news = crimeincident DF. Orop Duplico	utes O
newof, shows	181111
1. 11/2	
Finding name of a specific column	using index.

AMAL

P	OPU
To	picPage No. 34
1	7 mith Collenn
ے ۔	ScrimeincidentNC = VrimeinaidentDF, mithlohenn (
	'Incidn+Nem', crimeinciden+DF, Incidn+Nem), unth/dun
	C'category', vrimeincident DF. Category), mitholienn C'
	'Resolution', vimeincident OF. Resolution.
	orimeincidenta c. shows
	1 5 5 8 5 1 1. N 3 15 1 16 1 17 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
	Registering table as a template, and applying
	Registoring table as a template, and applying SOL queries to it.
	8 create Or Replace Temp View (table)
	* create Or Replace Temp View (table) * crênce incident NC. create Ox Replace Temp View ("crêne_incident
	table")
	newDF = sqlontext.sql ("select & from vine-incident
	table WHERE REsolution LIKE \ 'ARREY-\',")
	newof, show()
10	Filtering Oberations
	subset DF = orine incident NC. filter (wine incidents NC.
	Filtering Oberations Subset DF = Orime incident NC. filter (winne incidents NC. Resolution == "(NON 1=")")
	subset of show (10)
20	Eiltering through multiple conditions. Crêmeincidents NC. filter (crêmeincidents NC. Resorbetion
-	crimeincidents NC. filter (vimeincidents NC, Resoulution
	== LOCATIED?). filter (crimeincidents NC. (ategory == ASIANLT?)
	orimeinciclents NC, show C
	Teacher's Signature

POPU	
Date	
Paga No. 35	
- suime insidentif around some	
Category, vine incidents	F.
2) borouply -> vimeincident DF. groupsy ([(ategory', vrime incidents	
22 Applying Aggregations In alien	
23 Applying Aggregation functions after grouping. - Duineincident DF. groupsy (), ang (), show ()	
(a) Show()	
Aggregation using volument Resolution? = rewor = crimeincidentor, group By (crimeincidentor, Resolution) rewor. court(), snow()	
= remor = vrimeincidentof group By (syin associate to prose	
hew DF. count (), snow()	Wha
24 Display 15 elements of dataframe. - Simerici dent DF, show (15)	
- orlineinci dent DF, show (15)	
25 Writing the datagrame as Parquet file. Solmeincident DF, write, parquet (91 dbfs:/Filestone/taly	
-> Orbnemaident DF, write, Barquet (90 dbfs:/ Filestone / taly	les
(crime incidents, parquet")	
The Barquet file will be stored by default in HE - new DF - spark, read parquet ("dbfs:/Filestore/tables!) Ei
-> newDF = spark, read parquet ("dbfs: / Filestore / tables !	
vime incidents, parquet, infersehena = "true",	
header = "true")	
new DF, show()	
Teacher's Signature	

Outputs:

Out[4]: [Row (TheidntNiem = 50436212, Category = "ASSAULT",
Descript = 'BATTERY', Dayofweek = ! wednesday', Parte = !04/20/2025

12:00 AM', Time = '04:00', Polaistonict = 'MISSEON', RROULTION = 'NOW

Addrep = 1,8th ST/CASTROST '1)]

Out(8] crimeincidentDF, countly

OUT (9): Row (Incidn#Num = 50436=12) lategory = 1451ault,

Description='BATTERY', Dayof week = Wednesday', Bosto = 104/20/2005,

12:00:00 AM', Time = 04:00', PADistance = 'MISSION', Resolution='NONI

Addreps = '18Th ST/CASTRO ST', X = -126, U350, Y = 32, 760882,

Weatin = '(32,2609), -122, U350...)")

Teacher's Signature

Out[10]: Scort:

1- Incidn+Num: integer (nullable = +rue)

1- Category: storing (nullable = +rue)

1- Descript: storing (nullable = +rue)

1- Day (week: storing (nullable = +rue)

1- Date: Storing (nullable = +rue)

1- Time: Storing (nullable = +rue)

1- Poloistrict: storing (nullable = +rue)

1- Resolution: storing (nullable = +rue)

1- Address: storing (nullable = +rue)

1- X: double (nullable = +rue)

1- Y: double (nullable = +rue)

1- Uccation: storing (nullable = +rue)

	0000113:		()		en diagonalis a -	
	Summery	Incidnen	Category	Durning	Dandweek par	1
	count	1888567	188855	7	885620 188116	1
	mean	9.27355650.	MUL.		have hem	The second second
-	stdder	4.025831484	nuu		here new	The State of
	min	- 3949	ARSOM	MILLA	Buiday 101/01/25	
- Charles and a	marc.	991282127.	WEAPON LAWS	Cours	wednesday 11-61/6-	,
H						

OUT (12)	is the following the	Branda C. Carrella (A. 1997)
IncidentNum	Category	Devipt
50436712	ASSAULT	BATTERY
80049028	LARCENY/THEFT	GURANOTHEFT AUTO
130366639	ASSAULT	AUPRAVATED ASSAUCT

Teacher's Signature ..

OUT[147: Coverelation = 0.556556135460178

Out[16]: OrimeincidentDF[4] (olumn < Date >

OU+[18]:

Incidnt Num	Category	Descript	Dewolnkok	Date.	This
10000000	ASSAVIST	Am LA RAMINEN	C MACHALLE I	melante	Mula
30810835	DEINING	DRIVENH	TUESDAY	7/8/3	14:00
1	ict Res		DESIGNATION OF THE PROPERTY OF	X	Y+ Wation
INGLE	SIDE APRE	THOOKED =	DBLOCK OF STON	-122.M	M2 32-72 132 21
SOVIA	ERN ARRE	ST, DOOKED	HASON ST	155.NV	28 31.21 (32.)1.

Out (19).

	The state of the s	PRINCIPO (SERVICE DE LA CONTRACTOR DE L						-	
Oncidnt Nun	~ Category	Du	utilat	Damer	reek	Dat	L. V	Jame	Podliss
50436712	ASIAVLT	BA-	TTERY	Weeler	scon l	4/201	Carrier Carrier	0 0:0 p	PA PPE IVE
86064008	PRINEY			SWY		1 1		18:00	PARK
resolution.	ddress		X	Server Proposition Server Serv	J	The second second	o de overeg	ocation	The state of the s
NONE	18th ST / CASTRO S	J	-122.14	100	300	1 A	,	12.7/12	
NONE	1100 Block of						· A	2 3626	

out [21]:

Category	Codegory	Court
FRAUD	FRANS	36004
SULCIDE	SUICIDE	
LIQUOR LAWS	LIQUOR LAWA	3860
SECONDARY CODES	SECONDARY CODES	21636

Teacher's Signature

Lemetion.	The second secon
out(22]: Aggregation function	and (Y)
and (Incidn Nem) and (X)	1 7 7 9 5 0 /
9:27955637558 -122.42719103	34.741 +3 3500

OUT(25): Resolution	count	
JUVENILE BOOKED	11789	
EXCEPTIONAL CLEAR.	3699	
ARREST, BOOKED	NA 2 8/2 7	a print pri-pris
CLEARED - CONTACT	2	
, 3%	in the constitution	And it is it granges as in

part dels:/filestore/tables/come_incidents.parquet abready exists.

and the first of the second of the second of