

RiskFactor Analysis

▶

⌵

📖

✍

🗑

📄

⬇

⌛

?

⚙

default ▾

%spark

val hiveContext = new org.apache.spark.sql.hive.HiveContext(sc);

hiveContext.sql("SHOW tables").show()

hiveContext: org.apache.spark.sql.hive.HiveContext = org.apache.spark.sql.hive.HiveContext@78bf7bf9

tableName	isTemporary
avgmileage	false
drivermileage	false
emp_salary	false
geolocation	false
sample_07	false
sample_08	false
truckmileage	false
trucks	false

Took 3 seconds

FINISHED ▶ ⌵ 📖 ⚙

hiveContext.sql("DESCRIBE drivermileage").show()

col_name	data_type	comment
driverid	string	null
totmiles	bigint	null

Took 4 seconds

FINISHED ▶ ⌵ 📖 ⚙

hiveContext.sql("select * from drivermileage limit 10").show()

driverid	totmiles
A1	628507
A10	675377
A100	634338
A11	652452
A12	668241
A13	654319
A14	649645
A15	651036
A16	638086
A17	663821

Took 0 seconds

FINISHED ▶ ⌵ 📖 ⚙

09/06/2019

sandbox-hdp.hortonworks.com:9995/#!/notebook/2ECKW1546

RISK Factor Analysis

Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled | Untitled

Zeppelin

Notebook

Interpreter

Configuration

Search in your notebooks

FINISHED

Connected

RiskFactor Analysis

0

default

truckid	driverid	event	latitude	longitude	city	state	velocity	event_ind	idling_ind
A54	A54	normal	38.440467	-122.714431	Santa Rosa	California	17	0	0
A20	A20	normal	36.977173	-121.899402	Aptos	California	27	0	0
A40	A40	overspeed	37.957702	-121.29078	Stockton	California	77	1	0
A31	A31	normal	39.409608	-123.355566	Willits	California	22	0	0
A71	A71	normal	33.683947	-117.794694	Irvine	California	43	0	0
A50	A50	normal	38.40765	-122.947713	Occidental	California	0	0	1
A51	A51	normal	37.639097	-120.996878	Modesto	California	0	0	1
A19	A19	normal	37.962146	-122.345526	San Pablo	California	0	0	1
A77	A77	normal	37.962146	-122.345526	San Pablo	California	25	0	0

Took 1 seconds

FINISHED

```
val driver_occurance = hiveContext.sql("select driverid, count(driverid) occurrence from geolocation where event !='normal' group by driverid")

driver_occurance.registerTempTable("driver_occurrences")
hiveContext.sql("SHOW tables").show()
```

driver_occurance: org.apache.spark.sql.DataFrame = [driverid: string, occurrence: bigint]

tableName	isTemporary
driver_occurrences	true
avgmileage	false
drivermileage	false
emp_salary	false
geolocation	false
sample_07	false
sample_08	false
truckmileage	false
trucks	false

Took 0 seconds

FINISHED

```
val risk_factor = hiveContext.sql(" SELECT a.driverid,a.occurance,b.totmiles,b.totmiles/a.occurance riskfactor from driver_occurrences a JOIN drivermileage b where a.driverid = b.driverid")
risk_factor.saveAsTable("risk_factor")
hiveContext.sql("SHOW tables").show()
```

risk_factor: org.apache.spark.sql.DataFrame = [driverid: string, occurrence: bigint, totmiles: bigint, riskfactor: double]

warning: there were 1 deprecation warning(s); re-run with -deprecation for details

tableName	isTemporary
driver_occurrences	true
avgmileage	false
drivermileage	false
emp_salary	false
geolocation	false
risk_factor	false

Took 11 seconds

```
hiveContext.sql("SELECT * from risk_factor limit 15").show()
```

FINISHED ▶ ⌵ 📖 ⚙️

driverid	occurance	totmiles	riskfactor
A18	6	656474	109412.33333333333
A7	5	653787	130757.4
A8	2	653991	326995.5
A9	6	665456	110909.33333333333
A80	5	681115	136223.0
A81	6	645273	107545.5
A82	2	631722	315861.0
A83	3	653172	217724.0
A84	6	651620	108603.33333333333
A85	5	635947	127189.4
A86	7	629558	89936.85714285714
A87	6	665023	110837.16666666667
A88	5	665021	133004.2
A89	2	652530	326265.0
A20	5	631496	126299.2

Took 13 seconds

```
risk_factor.rdd.coalesce(1).saveAsTextFile("/tmp/data/risk_factor")
```

FINISHED    

Took 1 seconds (outdated)

```
hiveContext.sql("SHOW tables").show()
```

FINISHED ▶ ⌵ 📖 ⚙️

tableName	isTemporary
driver_occurrences	true
avgmileage	false
drivermileage	false
emp_salary	false
geolocation	false
risk_factor	false
sample_07	false
sample_08	false
truckmileage	false
trucks	false

Took 0 seconds

```
%sql
SELECT a.driverid, a.riskfactor, b.city, b.state
```

FINISHED ▶ 🔍 📖 ⚙️

FROM risk_factor a, geolocation b where a.driverid=b.driverid

Zeppelin

Notebook

Interpreter

Configuration

settings

driverid

riskfactor

city

state

RiskFactor Analysis

driverid

riskfactor

city

state

xAxis

driverid

yAxis

riskfactor

group

city

size

Apple Valley

Willits

San Quentin

Kneeland

Stockton

Ukiah

Homeland

La Palma

Santa Rosa

Redding

San Francisco

Ojai

Isleton

Palo Cedro

Knights Landing

La Quinta

Martinez

San Fernando

Knightsen

Glendora

San Pablo

Santa Maria

Marysville

Antelope

San Diego

Napa

Lodi

Klamath

Palmdale

Santa Monica

Irvine

Bakersfield

Mariposa

Modesto

Markleeville

Cloverdale

Palo Alto

La Puente

Santa Paula

Jacumba

San Dimas

Goleta

Aptos

Arbuckle

Occidental

326995.5

200000

5714285714

A18

A7

A8

A9

A80

A81

A82

A83

A84

A85

A86

A87

A88

Results are limited by 1000.

Took 1 seconds

READY

sandbox-hdp.hortonworks.com:9995/#!/notebook/2ECKW1546

4/4