## **Filtering Data**

Run

%pyspark # Read in data from S3 Buckets from pyspark import SparkFiles				
url ="https://s3.amazonaws.com/dataviz-curriculum/dspark.sparkContext.addFile(url) df = spark.read.csv(SparkFiles.get("wine.csv"), sep df.show()	)=",", h		=True)	10101
null! Tinta de Toro Bodega Carmen Rod	331	1101	погенети эритт	10101
USIThis re-named vin   Silicel alley   Pinot Noir   Bergstṛm	951	651	Oregon	Chehalem Mountains Willamette V
USIThe producer sour Gap's Crown Vineyard  onomal Pinot Noir  Blue Farm	951	601	Californial	Sonoma Coastl S
Italy Elegance, complex  Ronco della Chiesa  null  Friulano  Borgo del Tiglio	951	801	Northeastern Italy	Colliol
USIFrom 18-year-old Estate Vineyard W  alley  Pinot Noir Patricia Green Ce	951	481	Oregon	Ribbon Ridge Willamette V
USIA standout even il Weber Vineyardlalleyl Pinot NoirlPatricia Green Cel	951	481	Oregon	Dundee Hills Willamette V
France This wine is in p Chteau Montus Pr  null  Tannat  Vignobles Brumont	951	901	Southwest Francel	Madiranl
USIWith its sophisti   Grace Vineyard alley   Pinot Noir   Domaine Serene	951	1851	Oregon	Dundee Hills Willamette V
US First made in 200  Sigrid  alley  Chardonnay  Bergstṛm	951	901	0regon	Willamette Valley Willamette V
USIThis blockbuster,  Rainin Vineyard  Napa Cabernet Sauvignon  Hall	951	3251	CaliforniaID	iamond Mountain
Spain Nicely oaked blac 6 Aos Reserva Pr  null  Tempranillo  Valduero	951	801	Northern Spainl	Ribera del Duerol
France Coming from a sev  Le Pigeonnier  null  Malbec  Chteau Lagrzette	951	2901	Southwest Francel	Started Caholusp V (\$\sqrt{2}\) &
	~- '	·		· · ·

|-- country: string (nullable = true)
|-- description: string (nullable = true)
|-- designation: string (nullable = true)
|-- points: string (nullable = true)
|-- price: string (nullable = true)
|-- province: string (nullable = true)
|-- region\_1: string (nullable = true)
|-- region\_2: string (nullable = true)
|-- variety: string (nullable = true)
|-- winery: string (nullable = true)

```
%pyspark (/U4G66226D/spaces)
# Order a DataFrame by ascending values
df.orderBy(df["points"].asc()).show(5)
+----+
----+
                          designation|points|price|province|
Icountry
            description
                                                        reaion_1|
                                                                   region_2|variety|
   wineryl
+----+
                               null| null| null|
                                              nullI
                                                           nullI
                                                                     nullI
                                                                           nullI
   USIThis shows a deft...l
    nulll
                               null| null| null|
                                              nullI
                                                                           nullI
   USIThe strong aroma ...l
                                                           nulll
                                                                     nulli
    nullI
| Italy|"This nicely stru...| bitter almond an...| null|
                                               24|Northeastern Italy|Venezia Giulia|
                                         88 I
                                                                           nullIP
inot Grigiol
| Italy|This offers gener...|
                               null| null| null|
                                              nullI
                                                           nullI
                                                                           nullI
                                                                     nulll
    nullI
    USI
                Ripe I
                               null| null| null|
                                              nulli
                                                           nullI
                                                                     nullI
                                                                           nullI
    nullI
+-----+
----+
only showing top 5 rows
%pyspark
# Import functions
from pyspark.sql.functions import avg
df.select(avg("points")).show()
+----+
    avg(points)|
+----+
187.888341053831431
+----+
```

Run

%pyspark (/U4G66226D/spaces) # Using SQL df.filter("price<20").show()</pre> USTA SHIOKY SCETTE UTIO...T JIIIGEE LOC DECTION CULLIUITILUI MEHADELIO COULTEY IMEHADELIO/ LA ke Co...l Chardonnay I Parduccil | Portugal|This family-owned...| 10 I nulll Muros de Vinhal 86 I Dourol nulll Portuguese Redl Ouinta do Portall 14|Southwest France| Cahorsl FrancelThis is a smooth,...|Chteau Beauvilla...| 86 I nulll Malbec-Merlot1 Rigall USIThis is an easygo...I nullI 861 18 I Californial Californial Californ ia Otherl The Naked Grapel Chardonnay France|Closer to Bordeau...| Piamentuml 15|Southwest France| 86 I Buzetl Merlot-Malbect Georges Vigourouxl nulll FrancelThis is a blend o...l Pigmentuml 10|Southwest France| 861 Ctes de Gascognel null|Uani Blanc-Colombard| Georges Vigourouxl USIAromas of ripe (l...|Tudor Hills Vineyard| 17 l Washington Yakima Valleyl Columbi 86 I a Valley Pinot Grigiol Martinez & Martinezl USIStrong wood smoke... nulll 861 121 Californial Lodil Centra Cabernet Francl l Valley! Ironstonel USIThis medium-bodie...I nulll 86 I 101 Californial Californial Californ ia Otherl Chardonnay Leapina Horsel USIThis wine is dry ... 861 131 Californial Californial Californ nulll ia Otherl White Blendl Kitchen Sinkl | Portugal|This state-owned ...|Companhia das Lez...| 121 nulll 861 Tejol Portuguese Redl Wines & Winemakersl | | ArgentinalThis copper-tinte...|Terroir **©**\_nico Pi...| 12|Mendoza Province| Tupungatol 861 nulll Zorzall Rosl 15|Mendoza Province| Valle de Ucol Reserval 861 Vialbal nulll Malbecl -----+-----+

only showing top 20 rows

...

```
%pyspark (/U4G66226D/spaces)
# Filter by price on certain columns
df.filter("price<20").select(['points','country', 'winery','price']).show()
```

```
lpoints! country!
                                  winerylpricel
                            Villa Melnikl
     901 Bulgarial
                                             15 l
     90 I
             Spainl
                          Don Bernardinol
                                             17 I
                USI
     90 I
                                De Loachl
                                             181
                      Trinity Vineyards
     911
                USI
                                             191
     91| Portugal|Adega Cooperativa...|
                                             15 l
     861
                USI
                          Belle Ambiancel
                                             10 l
     861 Portugal! Adega de Cantanhede!
                                             12 l
     86 I
                USI
                                Parducci I
                                             131
                       Ouinta do Portall
                                             101
     86| Portugal|
     861
            Francel
                                             141
                                   Rigall
     861
                USI
                        The Naked Grapel
                                             181
                      Georges Vigourouxl
     861
                                             15 l
            Francel
                      Georges Vigourouxl
     861
            Francel
                                             101
     861
                USI Martinez & MartinezI
                                             17 I
     861
                USI
                                             121
                               Ironstonel
     861
                USI
                           Leaping Horsel
                                             10 l
     861
                USI
                           Kitchen Sinkl
                                             13 l
     86| Portugal|
                     Wines & Winemakers
                                             121
     86|Argentinal
                                  Zorzall
                                             121
                                 Vialbal
                                            15 l
     86|Argentinal
only showing top 20 rows
```

## **Using Python Comparison Operators**

%pyspark (/U4G66226D/spaces)
# Same results only this time using python
df.filter(df["price"] < 200).show()

	Direct Maint Paractum	LCCI JJ	1 051	or egonit	CHEHALEIII MOAITCALIISTW	LLLUMECCE VAL
	Pinot Noirl Bergstrml			6.7.6		-
Ι.	USIThe producer sour Gap's Crown Viney	ardl 95/	601	California	Sonoma Coastl	Son
oma l						
	taly Elegance, complex  Ronco della Chi	esal 95.	801	Northeastern Italy	Collio	n
ullI	Friulanol Borgo del Tigliol					
I	USIFrom 18-year-old Estate Vineyard W	VI 95	1 481	Oregonl	Ribbon Ridge W	illamette Val
leyl	Pinot Noir Patricia Green Ce					
I	USIA standout even i  Weber Viney	ardl 95/	1 481	Oregon	Dundee Hills W	illamette Val
leyl	Pinot Noir Patricia Green Ce					
l Fro	ance This wine is in p Cḫteau Montus Pr.	951	901	Southwest Francel	Madiranl	nu
111	Tannat  Vignobles Brumont					
I	USIWith its sophisti  Grace Viney	ardl 95/	1851	Oregon	Dundee Hills W	illamette Val
leyl	Pinot Noirl Domaine Serenel					
I	USIFirst made in 200l Sig	gridl 95	901	Oregon	Willamette Valley W	illamette Val
leyl	Chardonnayl Bergstṛml					
l Sr	pain Nicely oaked blac 6 Aos Reserva Pr.	951	80 I	Northern Spainl	Ribera del Duerol	nu
lli Tempranilloi Valdueroi						
I	USIThis fresh and li Gap's Crown Viney	ardl 95/	I 751	Californial	Sonoma Coastl	Son
oma l	Pinot Noirl Gary Farrell					
I	USIHeitz has made th  Grignol	inol 95	1 241	Californial	Napa Valleyl	N
apa l	Rosl Heitzl				,	
l Sr	pain Alluring, complex Prado Enea Gran R	RI 95	l 791	Northern Spainl	Riojal	n
	Tempranillo Blendl Mugal			·	•	
ı		arel 95	l 601	Californial	Edna Valleyl	Central Co
astl	Chardonnayl Center of Effortl					
	+	+	++	+	+-	
•	· +	•	•	·	•	
•	•					

only showing top 20 rows

Started Juno ✓ ﴿﴾ إ

•••

%pyspark (/U4G66226D/spaces) df.filter( (df["price"] < 200) | (df['points'] > 80) ).show() T Spatificasi cedary brack... rear odorani #\_fireo .... JJI TTOI 10101 nulli Tinta de ToroiBodega Carmen Rod...i 951 65 I USIThis re-named vin... Silicel Oregon| Chehalem Mountains|Willamette V Pinot Noirl alleyl Bergstrml USIThe producer sour...|Gap's Crown Vineyard| 951 60 I Californial Sonoma Coastl S Pinot Noirl Blue Farml onomal | Italy|Elegance, complex...| Ronco della Chiesa| Colliol 951 80|Northeastern Italy| Friulanol Borgo del Tigliol nulll USIFrom 18-year-old ...|Estate Vineyard W...| 951 481 Oregon l Ribbon RidgelWillamette V alleyl Pinot NoirlPatricia Green Ce...I Weber Vineyardl 951 Dundee Hills|Willamette V USIA standout even i...l 481 Oregon l Pinot NoirlPatricia Green Ce...I alleyl | France|This wine is in p...|Chteau Montus Pr...| 951 901 Southwest Francel Madiranl Tannatl Vignobles Brumontl nulll USIWith its sophisti... Grace Vineyard 951 185 l Oregon l Dundee Hills|Willamette V alleyl Domaine Serenel Pinot Noirl USIFirst made in 200... Sigridl 951 901 Oregon l Willamette Valley|Willamette V alley Chardonnay Bergstrml USIThis blockbuster,... Rainin Vineyard 325 I California|Diamond Mountain ...| 951 NapalCabernet Sauvignon! HallI | Spain|Nicely oaked blac...|6 Aos Reserva Pr...| 951 80 I Northern Spainl Ribera del Duerol Tempranillol Valduerol nullI I FrancelComing from a sev...l Le Pigeonnierl 2901 Southwest Francel 951 Cahors nullI Malbecl Chteau Lagrzettel USIThis fresh and li...|Gap's Crown Vineyard| 951 75 l Californial Sonoma Coastl S Gary Farrell! onomal Pinot Noirl 

only showing top 20 rows

%pyspark (/U4G66226D/spaces)

df.filter(df["country"] == "US").show()

description region\_1| lcountryl designation|points|price| province| region\_21 varietyl wineryl -----+ USIThis tremendous 1... | Martha's Vineyard 961 235|Californial Napa Valley Napa I Ca bernet Sauvignon! Heitzl USIMac Watson honors...|Special Selected ...| 961 90|Californial Knights Valley Sonoma I Sauvignon Blancl Macauleyl USIThis spent 20 mon... 65 l Willamette Valley|Willamette Valley| Reservel 961 Oregon l Pinot Noirl Ponzil USIThis re-named vin... Silicel 951 65 I Chehalem Mountains | Willamette Valley | Oregonl Pinot Noirl Bergstrml USIThe producer sour... | Gap's Crown Vineyard| 60|California| Sonoma Coastl 951 Sonomal Pinot Noirl Blue Farm! USIFrom 18-year-old ...|Estate Vineyard W...| 951 Ribbon RidgelWillamette Valley 481 Oregon l Pinot NoirlPatricia Green Ce...I USIA standout even i... Weber Vineyard 951 481 Oregon l Dundee Hills|Willamette Valley| Pinot NoirlPatricia Green Ce...l USIWith its sophisti... Grace Vineyard 1851 Oregon1 Dundee Hills|Willamette Valley| 951 Pinot Noirl Domaine Serenel USIFirst made in 200... Sigridl 901 Oregon1 Willamette Valley|Willamette Valley| 951 Chardonnayl Bergstrml USIThis blockbuster,... Rainin Vineyardl 325|California|Diamond Mountain ...| Napa I Ca 951 bernet Sauvignon! HallI USIThis fresh and li...|Gap's Crown Vineyard| 75|Californial Sonoma Coastl Sonomal 951 Pinot Noirl Gary Farrell US|Heitz has made th...| Grignolinol 951 24|Californial Napa Valley Napal

