: [pdf.head(5) [Row(ItemId='4', ItemName='C Row(ItemId='5', ItemName='Pe Row(ItemId='6', ItemName='Pe Row(ItemId='7', ItemName='De Row(ItemId='8', ItemName=Nor pdf.select("ItemId", "ItemName ItemId ItemName ItemCost Qe Chock 65.76 35 Pencil 45.65 nui 61 Pen 76.87 36	hock', ItemCost='65.76', Sencil', ItemCost='45.65', en', ItemCost='76.87', Suguster', ItemCost='54', Sugne, ItemCost=None, Supplieme", "ItemCost", "Qty").show+ ty + 34 11	Supplier='X', Grade=' Supplier='Y', Grade= oplier='X', Grade='A' oplier='Y', Grade='C' er=None, Grade=None,	='B', Qty=N ', Qty='23' ', Qty=None	one),),),		
. [6 Pen 76.87 2 7 Duster 54 nul 8 null null 3 9 book 53 6 10 null null nul pdf.dtypes [('ItemId', 'int'), ('ItemName', 'string'), ('ItemCost', 'double'), ('Supplier', 'string'), ('Grade', 'string'), ('Qty', 'int')] pdf.printSchema()	11 34 65 11					
:	root ItemId: string (nullable ItemName: string (nullable ItemCost: string (nullable Supplier: string (nullable Grade: string (nullable Qty: string (nullable Qty: string (nullable = Podf1=pdf.withColumn("Bill",ppdf1.show()	<pre>ble = true) ble = true) ble = true) = true) true) odf['ItemCost']*pdf['Qty'] +</pre>	Bill + 2235.84 null				
+	7 Duster 54.0 8 null null 9 book 53.0 10 null null pdf.withColumnRenamed('Qty', territory territory pdf.withColumnRenamed('Qty', territory pdf.withColumnRenamed('Qty',	Y C null null null 34 null null 65 null C null ++ pplier Grade Quantity	null null 3445.0 null				
+	9 book 53.0 10 null null # Drop Column pdf1.drop('Bill').show()	null null 65 null C null + pplier Grade Qty X A 34 Y B null X A 23 Y C null null null 34 null null 65					
+	pdf1.na.drop(how='any').show the state of t	pplier Grade Qty +++	Bill + 2235.84 0000000002 +				
. +	ItemId ItemName ItemCost Su	X A 34 Y B null X A 23 1768.016 Y C null null null 34 null null 65 null C null	null null 3445.0 null				
+	6 Pen 76.87 7 Duster 54.0 8 null null 9 Book 53.0 10 null null null hull	X A 23 1768.016 Y C null Z null 34 null null 65 null null null Show() Show()	0000000000002 null null 3445.0 null + de Qty	235.84			
+	10 Not Provided null 		C null ++ Bill + 2235.84 345.0 9000000002 345.0 345.0 3445.0 345.0	null			
+		X A 34 Y B null X A 23 Y C null Z null 34 null null 65 null null null	csv',inferSchema=Tru	ie)			
+ + + + + + + + + + + + + + + + + + + +	# minimun 3 nonnull values # pdf.na.drop(how='any',thresh	x A 34 Y B null X A 23 Y C null null null 65					
:	4 Chock 65.76 5 Pencil 45.65 6 Pen 76.87 7 Duster 54.0 8 null null	pplier Grade Qty					
+	9 book 53.0 10 null 0.0 pdf.na.fill(0,['ItemCost','S]	Supplier']).show()					
+	First Name Gender Start Date	-+	3onus % Senior Manage 6.945 4.17 11.858 f null 1.389 10.125 f 10.012 null null 7.524	ement true true false true false true fue fue fue true true	Team		
:	Julie Female 10/26/1997 Brandon Male 12/1/1980 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010 Donna Female	1:08 AM 112807 1 11:40 PM 109831 9 7:13 AM 41426 6 6:09 AM 59414 9 5:56 AM 90370 6 7:45 PM 111737 1 10:27 AM 132940 9 3:48 AM 81014	17.492 5.831 f 14.543 1.256 f 7.369 f 6.414 f 19.082 f 1.894 f	False true False False False False	+onus %	Teaml	
	count 933 855 mean null null stddev null null min Aaron Female max Willie Male pdf.columns ['First Name', 'Gender', 'Start Date', 'Last Login Time', 'Salary', 'Bonus %', 'Senior Management', 'Team']	null null null null 1/1/2014 10:00 AM 9/9/2006 9:59 AM	90497.76339737108 16 32851.15727674897 5	0.207108870 5.538046342	335046 1.015 Business Deve		
: [n	Row(First Name='Thomas', General Row(First Name='Maria', General Row(First Name='Jerry', General Row(First Name='Larry', General Row(First Name='Larry', General Row(First Name='Larry', General Row(First Name='Larry', General Row(First Name=', 'Telephone Row(First Name)	nder='Male', Start Date='3 der='Female', Start Date='3 der='Male', Start Date='1/ der='Male', Start Date='1/ eam','Salary']).show()++ Team Salary ++ ting 97308	3/31/1996', Last Logi '4/23/1993', Last Logi /4/2005', Last Login	in Time='6: gin Time='1 Time='1:00	53 AM', Salary=61933 1:17 AM', Salary=Non PM', Salary=138705,	, Bonus %=4.17, e, Bonus %=11.85 Bonus %=None, S	, Senior Management=True, Tea Senior Management=True, Team= 8, Senior Management=False, T enior Management=True, Team=' enior Management=True, Team='
	Maria Fina Jerry Fina Larry Client Servi Dennis Le Ruby Proc null Fina Angela Engineer Frances Business Develope Louise I Julie Le Brandon Human Resour Gary Sa Kimberly Fina Lillian Proc Jeremy Human Resour Shawn Proc Diana Client Servi	egal 115163 duct 65476 ance null ring 95570 ment 139852 null 63241 egal 102508 rces 112807 ales 109831 ance 41426 duct 59414 rces 90370 duct 111737					
: [pdf.dtypes [('First Name', 'string'), ('Gender', 'string'), ('Start Date', 'string'), ('Last Login Time', 'string ('Salary', 'int'), ('Bonus %', 'double'), ('Senior Management', 'boole ('Team', 'string')] pdf.printSchema()	'),					
+	First Name: string (null Gender: string (nullable Start Date: string (null Last Login Time: string Salary: integer (nullable Bonus %: double (nullable Senior Management: boole Team: string (nullable: pdf.na.drop(how='any').show(e = true) lable = true) (nullable = true) le = true) le = true) ean (nullable = true) = true) () -+	30nus % Senior Manage 	ement + true false true	Team		
	Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Donna Female 7/22/2010 Matthew Male 9/5/1999 John Male 7/1/1999 Craig Male 2/27/2000 Terry Male 11/27/1981 Benjamin Male 1/26/2009 Christina Female 8/6/2009 Jean Female 12/18/1999 Jerry Male 1/10/2000 Rachel Female 2/16/2009 Linda Female 10/19/1981 Stephanie Female 9/13/1980	7:13 AM 41426 6:09 AM 59414 6:09 AM 59414 7:45 PM 111737 7:45 PM 111737 7:45 AM 81014 7:45 AM 97950 7:45 AM 37598 10:08 PM 79529 1:19 PM 118780 1:19 PM 118780 1:19 PM 118780 1:156 PM 95734 1:156 PM 57427 1:52 AM 36844	14.543 1.256 7.369 6.414 1.894 13.645 13.873 7.757 13.464 7.008 9.096 16.18 19.096 12.599 9.557 5.574	true false false false false false false true true true false Busin false Busin true	Finance Product Product Product Product Product Marketing Client Services Legal Engineering ess Development Client Services ess Development Client Services ess Development Client Services ess Development		
+ - 	Thomas Male 3/31, Maria Female 4/23, Jerry Male 3/4, Larry Male 1/24, Dennis Male 4/18, Ruby Female Not Provided Female 8/12, Julie Female 10/26,	Date Last Login Time Sala +	ary Bonus % Senior Ma 308 6.945 933 4.17 111 11.858 705 null 111 1.389 163 10.125 176 10.012 111 null 1570 null 1570 null 152 7.524 1241 15.132 1508 12.637	true true false true false true false true false true	Team Marketing Not Provided Finance Finance Client Services Legal Product Finance Engineering usiness Development Not Provided Legal Human Resources		
: [Lillian Female 6/5, Jeremy Male 9/21, Shawn Male 12/7, Diana Female Not Prod Donna Female 7/22, Diana Female 7/22, Diana Female 7/22, Donly showing top 20 rows pdf.na.fill(7554).show()	7:13 AM 412 7:2016 6:09 AM 592 7:2010 5:56 AM 903 7:45 PM 1117 7:010 3:48 AM 810 7:45 PM 12:42 PM 97308 6 6:53 AM 61933 8 11:17 AM 7554	426 14.543 414 1.256 370 7.369 7.37 6.414 940 19.082 914 1.894 914 1.894 915 9	+ ement	+ Team		
	Larry Male 1/24/1998 Dennis Male 4/18/1989 Ruby Female 1/20/2019 Angela Female 1/20/2019 Angela Female 1/20/2019 Frances Female 1/21/1980 Julie Female 10/26/1999 Brandon Male 12/1/1980 Gary Male 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 1/2/7/1980 Diana Female 7/22/2010	8 4:47 PM 7554 7 1:35 AM 115163 1 null 65476 5 10:43 AM 7554 1 6:29 AM 95570 1 6:51 AM 139852 9 9:01 AM 63241 7 3:19 PM 102508 9 1:08 AM 112807 1 11:40 PM 109831 9 7:13 AM 41426 6 6:09 AM 59414 9 5:56 AM 90370 6 7:45 PM 111737 1 10:27 AM 132940	1.389 10.125 10.012 7554.0 7554.0 7.524 15.132 12.637 17.492 5.831 14.543 1.256 7.369 6.414 19.082 1.894	true false true null true true true true true false false false false false false	Client Services Legal Product Finance Engineering ess Development null Legal Human Resources Sales Finance Product Human Resources Product Client Services Product		
+	pdf.show() First Name Gender Start Date Douglas Male 8/6/1993 Thomas Male 3/31/1996 Maria Female 4/23/1993 Jerry Male 3/4/2009 Larry Male 1/24/1993 Dennis Male 4/18/1983 Ruby Female null null Female 7/20/2019 Angela Female null Frances Female null Louise Female 8/12/1986 Julie Female 10/26/1993 Brandon Male 12/1/1986	e Last Login Time Salary E -+	3onus % Senior Manage 6.945 4.17 11.858 f null 1.389 10.125 f 10.012 null null 7.524 15.132 12.637	ement true true false true false true fue true true true true true	Team		
:	pdf.na.drop(how='any',thresh	7:13 AM 41426 6 6:09 AM 59414 7:13 AM 41426 6 6:09 AM 59414 7:45 PM 111737 1 10:27 AM 132940 7:45 PM 132940 7:45	14.543 1.256 f 7.369 f 6.414 f 19.082 f 1.894 f 3onus % Senior Manage	False	+ Team		
	Jerry Male 3/4/2009 Larry Male 1/24/1998 Dennis Male 4/18/198 Frances Female null Louise Female 8/12/1980 Julie Female 10/26/1999 Brandon Male 12/1/1980 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female 7/22/2010 Lois null 4/22/1999 Matthew Male 9/5/1999 Joshua null 3/8/2012	8 4:47 PM null 7 1:35 AM 115163 1 6:51 AM 139852 9 9:01 AM 63241 7 3:19 PM 102508 9 1:08 AM 112807 1 11:40 PM 109831 9 7:13 AM 41426 6 6:09 AM 59414 9 5:56 AM 90370 6 7:45 PM 111737 1 10:27 AM 132940 9 3:48 AM 81014 7:18 PM 64714 2:12 AM 100612 1:58 AM 90816	1.389 10.125 17.524 15.132 12.637 17.492 5.831 14.543 1.256 17.369 16.414 19.082 1.894 1.894 13.645 18.816 18.816 17.816 18.816 17.816 18.816 18.816 18.816 17.524 18.816 17.524 18.816 17.524 18.816 17.524 18.816 18.816 17.524 18.816 17.524 18.816 17.524 18.816 17.524 18.816 17.524	false true Busin true true true false false false false false false false true false true	Finance Client Services		
+		-+	3onus % Senior Manage 6.945 4.17 null 10.125 10.012 null 7.524 15.132 12.637 17.492 5.831	ement true true true false true true true true true	Team +		
: + 	Lillian Female 6/5/2016 Jeremy Male 9/21/2016 Shawn Male 12/7/1986 Diana Female null Donna Female 7/22/2016 Lois null 4/22/1999 Matthew Male 9/5/1999 Joshua null 3/8/2012 Conly showing top 20 rows pdf.na.fill(0).show() First Name Gender Start Date	6 6:09 AM 59414 0 5:56 AM 90370 6 7:45 PM 111737 1 10:27 AM 132940 0 3:48 AM 81014 5 7:18 PM 64714 5 2:12 AM 100612 2 1:58 AM 90816	1.256 ff 7.369 ff 6.414 ff 19.082 ff 4.934 ff 13.645 ff 18.816 ff 30nus % Senior Manage 6.945	False False False False False true False true	Product Human Resources Product Client Services Product Legal Marketing Client Services + Team		
	Thomas Male 3/31/1996 Maria Female 4/23/1993 Jerry Male 3/4/2009 Larry Male 1/24/1998 Dennis Male 4/18/1983 Ruby Female null null Female 7/20/2019 Angela Female null Frances Female 8/12/1986 Julie Female 10/26/1993 Brandon Male 12/1/1986 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2016 Jeremy Male 9/21/2016 Shawn Male 12/7/1986 Diana Female 7/22/2016	11:17 AM 0 1:00 PM 138705 8	11.858 16	false true true true true true true true false false false false false false false	null Finance Finance Client Services Legal Product Finance Engineering ess Development null Legal Human Resources Finance Product Human Resources Product Client Services Product		
: + 	pdf.na.fill(0,['Bonus %','Sa pdf.na.fill(0,['Bonus %','Sa First Name Gender Start Date pouglas Male 8/6/1993 Thomas Male 3/31/1993 Maria Female 4/23/1993 Jerry Male 3/4/2003 Larry Male 1/24/1993 Dennis Male 4/18/1983 Ruby Female null null Female 7/20/2013 Angela Female null	-+	30nus % Senior Manage 6.945 4.17 11.858 f 0.0 1.389 10.125 f 10.012 0.0 0.0	ement true true false true false true fue false true	Team Team		
	Frances Female null Louise Female 8/12/1980 Julie Female 10/26/1999 Brandon Male 12/1/1980 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010	9:01 AM 63241 7 3:19 PM 102508 9 1:08 AM 112807 1 11:40 PM 109831 9 7:13 AM 41426 6 6:09 AM 59414 9 5:56 AM 90370 6 7:45 PM 111737 1 10:27 AM 132940 9 3:48 AM 81014	15.132 12.637 17.492 5.831 14.543 1.256 17.369 16.414 19.082 1.894 1	true true true false false false false	ess Development null Legal Human Resources Sales Finance Product Human Resources Product Client Services Product		
: : :	avg(Salary) 90497.76339737108 from pyspark.ml.feature impoimp=Imputer(inputCols=['SalatoutputCols=["{}] imp.fit(pdf).transform(pdf). First Name Gender Start Date	ary', 'Bonus %'], Imputed".format(s) for s i show() -+	Bonus % Senior Manage	+ ement		ry_Imputed Bo	nus %_Imputed +
 	Maria Female 4/23/1993 Jerry Male 3/4/2009 Larry Male 1/24/1993 Dennis Male 4/18/1983 Ruby Female null null Female 7/20/2019 Angela Female null Frances Female null Louise Female 8/12/1980 Julie Female 10/26/1993 Brandon Male 12/1/1980 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female 7/22/2010	10:43 AM null 1	null null 7.524 15.132 12.637 17.492 5.831 14.543 1.256 7.369 6.414 19.082 1.894	false true true true true true true true false false false false false false	Finance Finance Finance Client Services Legal Product Finance Engineering ess Development null Legal Human Resources Finance Product Human Resources Product Client Services Product	90497 115163 65476 90497 10.20 95570 10.20 139852 63241 102508 112807 109831 41426 59414 90370 111737 132940 81014	11.858 7108870967739
:	imp.fit(pdf).transform(pdf).	ary', 'Bonus %'], [mputed".format(s) for s i .show() -+	3onus % Senior Manage +	+ ement + true true [[] alse true		ry_Imputed Bonus	%_Imputed
 	Dennis Male 4/18/198 Ruby Female null null Female 7/20/2019 Angela Female null Frances Female null Louise Female 8/12/1980 Julie Female 10/26/1999 Brandon Male 12/1/1980 Gary Male null Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female 7/22/2010 Donna Female 7/22/2010	1:35 AM 115163 null 65476 10:43 AM null 6:29 AM 95570 11 6:51 AM 139852 9:01 AM 63241 7:3:19 PM 102508 9:1:08 AM 112807 11:40 PM 109831 7:13 AM 41426 6:09 AM 59414 9:01 5:56 AM 90370 6:01 7:45 PM 11737 11 10:27 AM 132940 11 3:48 AM 81014	10.125 ff 10.012 null null 7.524 15.132 12.637 17.492 5.831 ff 14.543 1.256 ff 7.369 ff 6.414 ff 19.082 ff	false true null true true Busin true true true false false false false false false	Legal Product Finance Engineering ess Development null Legal Human Resources Sales Finance Product Human Resources Product Client Services Product	115163 65476 86676 95570 139852 63241 102508 112807 109831 41426 59414 90370 111737 132940 81014	10.125 10.012 8.999 8.999 7.524 15.132 12.637 17.492 5.831 14.543 1.256 7.369 6.414 19.082 1.894
	First Name Gender Start Date Kathy Female 3/18/2006 null Female 2/23/2009 Rose Female 5/28/2019 Katherine Female 8/13/1996 Russell null 5/9/2009 Andrea Female 7/22/1999 Cynthia Female 7/12/2006 pdf.filter(pdf["Salary"] >=	-+	30nus % Senior Manage 16.991 1.825 5.63 f 18.912 f 3.533 f 13.707 7.864 f	ement true null false Human false false true Di false	Team Finance Sales Resources Finance Marketing stribution Product		
:	First Name Gender Start Date Kathy Female 3/18/2006 null Female 2/23/2009 Rose Female 5/28/2019 Katherine Female 8/13/1996 Russell null 5/9/2009 Andrea Female 7/22/1999 Cynthia Female 7/12/2006 pdf.filter(~(pdf["Salary"] > First Name Gender Start Date Douglas Male 8/6/1993	e Last Login Time Salary E -+	30nus % Senior Manage 16.991 1.825 5.63 18.912 3.533 13.707 7.864 6.945	ement true null false Human false false true Di false true true true true	Team Finance Sales Resources Finance Marketing stribution Product + Team Marketing		
		12:42 PM 97308 6 6:53 AM 61933 1:00 PM 138705 7 1:35 AM 115163 1 null 65476 1 6:29 AM 95570 1 6:51 AM 139852 9 9:01 AM 63241 7 3:19 PM 102508 0 1:08 AM 112807 1 11:40 PM 109831 9 7:13 AM 41426 6 6:09 AM 59414 0 5:56 AM 90370 7:45 PM 111737 10:27 AM 132940 3:48 AM 81014 7:18 PM 64714 5:1 2:12 AM 100612 2 1:58 AM 90816	6.945 4.17 null 10.125 10.012 null 7.524 15.132 12.637 17.492 5.831 14.543 1.256 7.369 6.414 19.082 1.894 4.934 13.645 18.816	true true true true true true true true Busin true true false false false false false false false true false true false true	Marketing null Finance Legal Product Engineering ess Development null Legal Human Resources Sales Finance Product Human Resources Product Client Services Legal Marketing Client Services		
	Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010 Lois null 4/22/1990 Matthew Male 9/5/1990 Joshua null 3/8/2012	1:58 AM 90816 -+	18.816 == 80000)).show() ====================================	true + ement true true true true true true true true	Client Services + Team + null Product null Finance Product Legal Client Services Marketing Legal		
	Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010 Lois null 4/22/1999 Matthew Male 9/5/1999 Joshua null 3/8/2012 Ponly showing top 20 rows pdf.filter((pdf["Salary"] >= First Name Gender Start Date First Name Gender Start Date Ruby Female 3/31/1990 Ruby Female 8/12/1980 Kimberly Female 1/14/1999 Lillian Female 6/5/2010 Lois null 4/22/1999 null Male 10/8/2012 Craig Male 2/27/2000 Benjamin Male 1/26/2009	1 null 65476 0 9:01 AM 63241 9 7:13 AM 41426 6 6:09 AM 59414 5 7:18 PM 64714 2 1:12 AM 37076 0 7:45 AM 37598 5 10:06 PM 79529	7.757 7.008	true true Busin true Busin true	Legal Client Services ess Development ess Development		
: + + + + + + + + + + + + + + + + + + +	Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010 Lois null 4/22/1999 Matthew Male 9/5/1999 Joshua null 3/8/2012 Shawing top 20 rows Donly showing	7:13 AM 41426 6 6:09 AM 59414 7:18 PM 64714 2 1:12 AM 37076 0 7:45 AM 37598 5 10:06 PM 79529 1 8:49 PM 57427 6 1:52 AM 36844 5 1:08 AM 66582 0 3:16 AM 73524 5 4:51 AM 66820 1 8:29 AM 41126 0 2:41 AM 49339 4 1:28 PM 40341 6 1:44 AM 64715 0 7:18 AM 72670 7 9:32 AM 66851	7.757 7.008 9.557 5.574 11.308 5.207 9.0 14.009 1.695 17.578 15.107 1.481 11.279 1	null true true true true false +	+ Team		
: + + + + + + + + + + + + + + + + + + +	Jeremy Male 9/21/2010 Shawn Male 12/7/1980 Diana Female null Donna Female 7/22/2010 Lois null 4/22/1999 Matthew Male 9/5/1999 Joshua null 3/8/2012 Showing top 20 rows Donly showing	7:13 AM 41426 6:09 AM 59414 7:18 PM 64714 1:12 AM 37076 7:45 AM 37598 10:06 PM 79529 11	7.757 7.008 9.557 5.574 11.308 5.207 9.0 14.009 1.695 17.578 15.107 1.481 11.279 6.945 4.17 10.012 null 15.132 14.543 1.256 7.369 1.894 4.934 18.816 13.873 18.576 7.757 7.008 12.752 19.096	null true true true true false true false false true true true true false false false true false	Human Resources Engineering Engineering + Team		