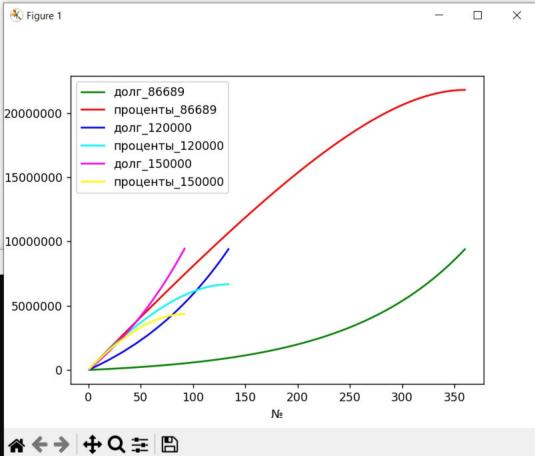
T. [Vind=: ine:			_
Python 3.8.0 Shell	_	×	
File Edit Shell Debug Options Window Help			
Python 3.8.0 (tags/v3.8.0:fa919fd, Oct 14 2019, 19:37:50) [MD64)] on win32 Type "help", "copyright", "credits" or "license()" for more		(AM	1
>>>			
=======		 	



```
d4.py - C:\spark\d4.py (3.8.0)
File Edit Format Run Options Window Help
 1 import pyspark, time, platform, sys, os
 2 from datetime import datetime
 3 from pyspark.sql.session import SparkSession
 4 from pyspark.sql.functions import col, lit, current timestamp
 5 import pandas as pd
 6 import matplotlib.pyplot as plt
 7 from sqlalchemy import inspect, create engine
 8 from pandas.io import sql
 9 import warnings, matplotlib
10 warnings.filterwarnings("ignore")
11 t0=time.time()
12 con=create engine("mysql://root:root@localhost/spark")
13 os.environ['PYSPARK PYTHON'] = sys.executable
14 os.environ['PYSPARK DRIVER PYTHON'] = sys.executable
15 spark=SparkSession.builder.appName("Hi").getOrCreate()
16
17 sql.execute("""drop table if exists spark.d4 1""",con)
18 sql.execute("""CREATE TABLE if not exists spark.d4 1 (
           'N' INT(10) NULL DEFAULT NULL,
19
20
            'Mecau' DATE NULL DEFAULT NULL,
21
            'Cymma nnatema' FLOAT NULL DEFAULT NULL,
22
           'Платеж по основному долгу' FLOAT NULL DEFAULT NULL,
23
            'Платеж по процентам' FLOAT NULL DEFAULT NULL,
24
            'OCTATOR GONFA' FLOAT NULL DEFAULT NULL,
25
            'npouenth' FLOAT NULL DEFAULT NULL,
26
            ' HOMP' FLOAT NULL DEFAULT NULL
27 )
28 COLLATE='utf8mb4 0900 ai ci'
29 ENGINE=InnoDB""", con)
30 from pyspark.sql.window import Window
31 from pyspark.sql.functions import sum as sum1
32 w = Window.partitionBy(lit(1)).orderBy("N").rowsBetween(Window.unboundedPreceding, Window.currentRow)
33 df1 = spark.read.format("com.crealytics.spark.excel")
            .option("sheetName", "Sheet1") \
34
            .option("useHeader", "false") \
36
            .option("treatEmptyValuesAsNulls", "false") \
37
            .option("inferSchema", "true").option("addColorColumns", "true")\
            .option("usePlainNumberFormat", "true")\
39
            .option("startColumn", 0)\
40
            .option("endColumn", 99)\
            .option("timestampFormat", "MM-dd-yyyy HH:mm:ss")\
41
42
            .option("maxRowsInMemory", 20)\
43
            .option("excerptSize", 10)\
            .option("header", "true") \
44
45
            .format("excel") \
46
            .load("/spark/d4 1.xlsx").limit(1000)\
47
            .withColumn("проценты", sum1(col("Платеж по процентам")).over(w))\
48
            .withColumn("долг", sum1(col("Платеж по основному долгу")).over(w))
49 df1.write.format("jdbc").option("url","jdbc:mysq1://localhost:3306/spark?user=root&password=root")\
50
            .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "d4 1")\
51
            .mode ("append") .save()
52 df2 = df1.toPandas()
53 # Get current axis
54 \, ax = plt.gca()
55 ax.ticklabel format(style='plain')
56 # bar plot
57 df2.plot(kind='line',
58
           x='N',
59
           у='долг',
           color='green', ax=ax)
60
61 df2.plot(kind='line',
62
           x='N'',
63
           у='проценты',
           color='red', ax=ax)
64
65
66 and avanuta ( III II dwan table if avieta anable de all III ann)
```

```
d4.py - C:\spark\d4.py (3.8.0)
File Edit Format Run Options Window Help
            color='red', ax=ax)
64
65
66 sql.execute("""drop table if exists spark.d4 2""",con)
67 sql.execute("""CREATE TABLE if not exists spark.d4 2 (
68
           'Nº' INT(10) NULL DEFAULT NULL,
69
            'Mecau' DATE NULL DEFAULT NULL,
70
            'Cymma nnatema' FLOAT NULL DEFAULT NULL,
71
           'Платеж по основному долгу' FLOAT NULL DEFAULT NULL,
72
           'Платеж по процентам' FLOAT NULL DEFAULT NULL,
73
            'OCTATOR GONDA' FLOAT NULL DEFAULT NULL,
74
            'npouentm' FLOAT NULL DEFAULT NULL,
75
            'долг' FLOAT NULL DEFAULT NULL
76)
77 COLLATE='utf8mb4 0900 ai ci'
78 ENGINE=InnoDB"", con)
79 from pyspark.sql.window import Window
80 from pyspark.sql.functions import sum as sum1
81 w = Window.partitionBy(lit(1)).orderBy("N").rowsBetween(Window.unboundedPreceding, Window.currentRow)
82 df3 = spark.read.format("com.crealytics.spark.excel")
83
           .option("sheetName", "Sheet1") \
84
            .option("useHeader", "false") \
            .option("treatEmptyValuesAsNulls", "false")\
86
            .option("inferSchema", "true").option("addColorColumns", "true")\
87
            .option("usePlainNumberFormat", "true") \
            .option("startColumn", 0)\
89
           .option("endColumn", 99)\
90
           .option("timestampFormat", "MM-dd-yyyy HH:mm:ss")\
 91
           .option("maxRowsInMemory", 20)\
92
            .option("excerptSize", 10)\
93
           .option("header", "true") \
94
            .format("excel") \
95
            .load("/spark/d4 2.xlsx").limit(1000)\
96
            .withColumn("проценты", sum1(col("Платеж по процентам")).over(w))\
97
            .withColumn("долг", sum1(col("Платеж по основному долгу")).over(w))
98 df3.write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")\
99
            .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "d4 2")\
100
            .mode ("append") .save ()
101 df4 = df3.toPandas()
102 ax.ticklabel format(style='plain')
103 # bar plot
104 df4.plot(kind='line',
105
           X= ' Nº '
106
           у='долг',
107
           color='blue', ax=ax)
108 df4.plot(kind='line',
109
           X = N_i
110
           у='проценты',
111
           color='cyan', ax=ax)
112
113 sql.execute("""drop table if exists spark.d4 3""",con)
114 sql.execute("""CREATE TABLE if not exists spark.d4 3 (
           'Nº' INT(10) NULL DEFAULT NULL,
115
116
           'Mecau' DATE NULL DEFAULT NULL,
117
           `Cymma платежа` FLOAT NULL DEFAULT NULL,
           'Платеж по основному долгу' FLOAT NULL DEFAULT NULL,
118
119
            'Платеж по процентам' FLOAT NULL DEFAULT NULL,
120
            'OCTATOR GONTA' FLOAT NULL DEFAULT NULL,
            'npouentm' FLOAT NULL DEFAULT NULL,
121
           ' HONT' FLOAT NULL DEFAULT NULL
122
123)
124 COLLATE='utf8mb4 0900 ai ci'
125 ENGINE=InnoDB""", con)
126 from pyspark.sql.window import Window
127 from pyspark.sql.functions import sum as sum1
128 w = Window.partitionBy(lit(1)).orderBy("№").rowsBetween(Window.unboundedPreceding, Window.currentRow)
```

```
d4.py - C:\spark\d4.py (3.8.0)
File Edit Format Run Options Window Help
TOUT HOLL PIOU
104 df4.plot(kind='line',
105
           x='N''
106
            у='долг',
107
           color='blue', ax=ax)
108 df4.plot(kind='line',
109
           x='N',
110
           у='проценты',
111
           color='cyan', ax=ax)
112
113 sql.execute("""drop table if exists spark.d4 3""",con)
114 sql.execute("""CREATE TABLE if not exists spark.d4_3 (
115
            'N' INT(10) NULL DEFAULT NULL,
116
            'Mecau' DATE NULL DEFAULT NULL,
117
            'Cymma nnatema' FLOAT NULL DEFAULT NULL,
118
            'Платеж по основному долгу' FLOAT NULL DEFAULT NULL,
            'Платеж по процентам' FLOAT NULL DEFAULT NULL,
119
120
            'OCTATOR GONFA' FLOAT NULL DEFAULT NULL,
121
            'npouentm' FLOAT NULL DEFAULT NULL,
122
            ' DONF' FLOAT NULL DEFAULT NULL
123)
124 COLLATE='utf8mb4 0900 ai ci'
125 ENGINE=InnoDB""", con)
126 from pyspark.sql.window import Window
127 from pyspark.sql.functions import sum as sum1
128 w = Window.partitionBy(lit(1)).orderBy("№").rowsBetween(Window.unboundedPreceding, Window.currentRow)
129 df5 = spark.read.format("com.crealytics.spark.excel")\
            .option("sheetName", "Sheet1") \
130
131
            .option("useHeader", "true")\
132
            .option("treatEmptyValuesAsNulls", "false") \
            .option("inferSchema", "true").option("addColorColumns", "true")\
133
134
            .option("usePlainNumberFormat", "true") \
135
            .option("startColumn", 0)\
136
            .option("endColumn", 99)\
137
            .option("timestampFormat", "MM-dd-yyyy HH:mm:ss")\
138
            .option("maxRowsInMemory", 20)\
139
            .option("excerptSize", 10)\
            .option("header", "true") \
140
141
            .format("excel") \
            .load("/spark/d4 3.xlsx").limit(1000)\
142
143
            .withColumn("проценты", sum1(col("Платеж по процентам")).over(w))\
144
            .withColumn("долг", sum1(col("Платеж по основному долгу")).over(w))
145 df5.write.format("jdbc").option("url","jdbc:mysql://localhost:3306/spark?user=root&password=root")
146
            .option("driver", "com.mysql.cj.jdbc.Driver").option("dbtable", "d4_3")\
147
            .mode ("append") .save()
148 df6 = df5.toPandas()
149 ax.ticklabel format(style='plain')
150 # bar plot
151 df6.plot(kind='line',
152
           x='N'',
153
           у='долг',
           color='magenta', ax=ax)
154
155 df6.plot(kind='line',
156
           X = 'N''
157
           у='проценты',
158
           color='yellow', ax=ax)
159
160 # show the plot
161 plt.legend(['долг 86689', 'проценты 86689', 'долг 120000', 'проценты 120000', 'долг 150000', 'проценты 150000'])
162 plt.show()
163
164 spark.stop()
165 t1=time.time()
166 print('finished', time.strftime('%H:%M:%S', time.gmtime(round(t1-t0))))
167
168
```

174 SELECT * FROM `spark`.'d4_3` LIMIT 1000; 175 SHOW CREATE TABLE `spark`.'d4_1`; 176 SELECT * FROM `spark`.'d4_1` LIMIT 1000;

176 SELECT * FROM `spark`.'d4_1` LIMIT 1000; 177 SHOW CREATE TABLE `spark`.'d4_2`; 178 SELECT * FROM `spark`.'d4_2` LIMIT 1000; Unnamed\spark\d4_3\ - HeidiSQL 12.7.0.6850 Файл Редактировать Поиск Запрос Инструменты Переход Помощь 🛴 Фильтр баз д 代 Фильтр табль 👚 📕 Unnamed 📒 База данных: spark 🟢 Таблица: d4_3 🔠 Данные 🕟 Запрос* 🗒 ∨ N Unnamed spark.d4_3: 92 строк (точно) information_sc... NΩ Месяц Сумма платежа Платеж по основному долгу Платеж по процентам Остаток долга проценты долг mysql 2024-06-25 150 000 65 605,5 84 394,5 9 397 710 84 394,5 65 605,5 2 2024-07-25 150 000 68 897.9 81 102.1 9 328 810 165 497 134 503 performance s... 3 2024-08-25 150 000 66 813 83 187 9 262 000 248 684 201 316 spark 1.9 MiB 2024-09-25 150 000 67 412,9 82 587,1 9 194 580 331 271 268 729 d4 48.0 KiB 5 2024-10-25 150 000 70 662,7 79 337,3 9 123 920 410 608 339 392 d4_1 48,0 KiB 6 2024-11-25 150 000 68 652,6 81 347,4 9 055 270 491 955 408 045 d4_2 16.0 KiB 2024-12-25 150 000 71 873,2 78 126,8 8 983 390 570 082 479 918 d4_3 16,0 KiB 8 2025-01-25 150 000 69 737,3 80 262,7 8 913 660 650 345 549 655 9 619 978 2025-02-25 150 000 70 322,6 79 677,4 8 843 330 730 022 s3_1 16,0 KiB 2025-03-25 10 10 150 000 78 605,2 71 394,8 8 764 730 801 417 698 583 s3_2 16.0 KiB 11 2025-04-25 770 246 11 150 000 71 663,4 78 336,6 8 693 060 879 754 s3_3 16.0 KiB 12 2025-05-25 954 939 845 061 12 150 000 74 814.8 75 185,2 8 618 250 s3_4 16,0 KiB 13 13 2025-06-25 150 000 72 982.1 77 017,9 8 545 270 1 031 960 918 043 54 48,0 KiB 14 14 2025-07-25 150 000 76 102,4 73 897,6 8 469 160 1 105 850 994 146 tabr 16,0 KiB 15 15 2025-08-25 150 000 74 324,3 75 675.7 8 394 840 1 181 530 1 068 470 16 2025-09-25 150 000 74 993.4 75 006,6 8 319 850 1 256 540 1 143 460 16 16,0 KiB tabr1 17 17 2025-10-25 150 000 78 066,3 71 933,7 8 241 780 1 328 470 1 221 530 16.0 KiB tabr2 2025-11-25 150 000 76 371,4 8 165 410 1 402 100 1 297 900 18 73 628.6 18 tabr3 48.0 KiB 19 19 2025-12-25 150 000 79 411,9 70 588,1 8 086 000 1 472 690 1 377 310 tabr4 48,0 KiB 20 2026-01-25 150 000 77 773,8 72 226,2 8 008 220 1 544 910 1 455 090 20 tasketl2a 1,5 MiB 21 2026-02-25 150 000 78 474 71 526 7 929 750 1 616 440 1 533 560 21 22 2026-03-25 150 000 86 034 7 843 720 22 63 966 1 680 410 1 619 590 > sys 23 23 2026-04-25 150 000 79 955 70 045 7 763 760 1 750 450 1 699 550 24 24 2026-05-25 150 000 82 911.1 67 088.9 7 680 850 1 817 540 1 782 460 25 2026-06-25 150 000 81 421,3 68 578,7 7 599 430 1 886 120 1 863 880 25 26 2026-07-25 150 000 84 342.9 65 657.1 7 515 090 1 951 780 1 948 220 26 27 27 2026-08-25 150 000 82 913,6 67 086,4 7 432 170 2 018 860 2 031 140 28 28 2026-09-25 150 000 83 660,1 66 339,9 7 348 510 2 085 200 2 114 800 29 2026-10-25 150 000 86 528,9 63 471,1 7 261 980 2 148 670 2 201 330 30 30 2026-11-25 150 000 85 192,2 64 807,8 7 176 790 2 213 480 2 286 520 31 31 2026-12-25 150 000 88 025 61 975 7 088 770 2 275 460 2 374 540 32 32 2027-01-25 150 000 86 751,7 63 248,3 7 002 010 2 338 700 2 461 300 33 2027-02-25 150 000 87 532,7 62 467,3 6 914 480 2 401 170 2 548 830 34 34 2027-03-25 150 000 94 289,7 55 710,3 6 820 190 2 456 880 2 643 120 35 35 2027-04-25 150 000 89 169,6 60 830,4 6 731 020 2 517 710 2 732 290 36 36 2027-05-25 150 000 91 908,7 58 091,3 6 639 110 2 575 800 2 824 200 37 37 2027-06-25 150 000 90 799,8 59 200,2 6 548 310 2 635 000 2 915 000 38 2027-07-25 150 000 93 500,5 56 499,5 6 454 810 2 691 500 3 008 500 38 39 39 2027-08-25 150 000 92 459 57 541 6 362 350 2 749 040 3 100 960 40 40 2027-09-25 150 000 93 291,4 56 708,6 6 269 060 2 805 750 3 194 250 41 41 2027-10-25 150 000 95 933,5 54 066,6 6 173 130 2 859 820 3 290 180 42 42 2027-11-25 150 000 94 994,9 55 005,1 6 078 140 2 914 820 3 385 180 43 43 2027-12-25 150 000 97 596,9 52 403,1 5 980 540 2 967 230 3 482 770 44 44 2028-01-25 150 000 96 846,1 53 153,9 5 883 690 3 020 380 3 579 620 45 45 2028-02-25 150 000 97 743,8 52 256,2 5 785 950 3 072 640 3 677 360 46 150 000 2028-03-25 101 936 48 063,9 5 684 010 3 120 700 3 779 300 20 525 5 ED 455 4 3 474 460 2 070 040 Фильт Регулярное выражение

177 SHOW CREATE TABLE `spark`.`d4_2`;
178 SELECT * FROM `spark`.`d4_2` LIMIT 1000;
179 SHOW CREATE TABLE `spark`.`d4_3`;
180 SELECT * FROM `spark`.`d4_3` LIMIT 1000;