My WSB My courses WSB Merito-DSW Digital Library Extranet Support >

ersytet eśląski DSW aw

Spark - introduction

Spark Core (RDD)

Spark - introduction - tutori...

Spark Core (RDD) - worksh...

SP03_w1_23_Spark-RDD-z...

Spark Core (RDD) - worksh...

Spark - introduction, RDD - ...

▼ 7. Spark SQL... Highlighted

Readings

Spark - DataFrame

Spark - DataFrame - works...

SP04_w1_24-Spark-DataFr...

Spark - DataFrame - works... Spark SQL - DataFrame - q...

✓ Project 2

Project description

Project notebook (template)

Przykład rozwiązania proje...

Datasets description

Project 2 - submission



Poniżej znajdą Państwo sugerowane wyniki jakie powinny być uzyskane w ramach poszczególnych zestawów.

W przypadku tych zestawów, które zawierają * w tytule, wyniki zostały potwierdzone.

W przypadku pozostałych zestawów uzyskanie innego wyniku nie oznacza, że jest on błędny. Należy w takim przypadku zgłosić się do prowadzącego celem weryfikacji rozwiązania i ewentualnej korekty sugerowanego wyniku na tej stronie.

Zestaw 1 - imdb-genres*

+	+		
decade	genre		actors
	+		+
	ocumentary	18	9
1890	News	5	9
1890	Sport	6	9
1900	Drama	22	46
1900	Comedy	10	17
1900	Biography	3	15
1910	Drama	5848	9624
1910	Comedy	1268	3433
1910	Crime	675	2104
1920	Drama	7060	11588
1920	Comedy	2196	4863
1920	Romance	1533	3534
1930	Drama	7872	10798
1930	Comedy	4179	6619
1930	Romance	2438	4452
1940	Drama	5834	9619
1940	Comedy	3449	5911
1940	Romance	1655	3752
1950	Drama	9450	14667
1950	Comedy	4503	8155
1950	Romance	2383	5337
1960	Drama	11992	21529
1960	Comedy	5678	11342
1960	Romance	3103	6695
1970	Drama	13577	27579
1970	Comedy	6573	15616
1970	Crime	2835	7690
1980	Drama	14590	31414
1980	Comedy	6873	17509
1980	Action	4605	9984
1990	Drama	14327	33326
1990	Comedy	7651	20121
1990	Action	5234	11349
2000	Drama	24542	65115
2000	Comedy	13461	40421
2000 D	ocumentary	20332	33972
2010		51949	
2010 D	ocumentary	53549	84988
2010	Comedy	26157	78062
2020	Drama	1349	3994
2020	Comedy	714	2357
2020	Horror	550	1623

Zestaw 2 - nyc-taxi*

			trip_distance					
2018-12 Manhattan							87599}, {08, 85825}, {15, 84043}]	
2018-12 Unknown	75,779	697,062					2279}, {19, 2030}, {18, 1966}]	
2018-12 Queens	264,815	5,389,044	1,525,387	1	[{27,	6310}, {21, 6108}, {28, 5987}]	
2018-11 Unknown	64,540	625,450	105,210	1	[{10,	1923}, {23, 1859}, {14, 1710}]	
2018-11 Manhattan	3,459,143	26,047,332	4,469,622	1	[{10,	87158}, {03, 85184}, {17, 84468}]	
2018-11 Queens	251,199	5,314,633	1,501,461	I	[{16,	6653}, {27, 5813}, {01, 5718}]	

Zestaw 3 - nyc-accidents*

1	street	borough	person_type :	injured	cilled
+	+	+	+	+	+
1	1 AVENUE	MANHATTAN	cyclist	416.0	1.0
I	BROADWAY	MANHATTAN	cyclist	468.0	1.0
1	2 AVENUE	MANHATTAN	cyclist	479.0	0.0
NORTHERN	BOULEVARD	QUEENS	motorist	1935.0	4.0
LINDEN	BOULEVARD	BROOKLYN	motorist	2409.0	5.0
ATLANT	IC AVENUE	BROOKLYN	motorist	2861.0	5.0
L	2 AVENUE	MANHATTAN	pedestrians	846.0	7.0
1	3 AVENUE	MANHATTAN	pedestrians	858.0	11.0
I	BROADWAY	MANHATTAN	pedestrians	1050.0	9.0
+	+		+		+

Zestaw 4 - imdb-persons*

+-	+	+	+
Ιp	rofession	primaryName	movies
+.	+		+
ı	actor	Luis Eduardo Motoa	3559
ı	actor	Ronit Roy	2602
ı	actor	Dilip Joshi	2385
ı	actress	Luz Stella Luengas	3636
ı	actress	Rohini Hattangadi	3240
ı	actress	Kavita Lad	3204
l	producer	Shobha Kapoor	11833
ı	producer	Ekta Kapoor	8826
ı	producer	Valentin Pimstein	6081
ı	writer	Tony Warren	6153
ı	writer	Delia Fiallo	6132
ı	writer	Sampurn Anand	5205

Zestaw 5 - netflix-prize-data

Zestaw 6 - fifa-players*

category	name	sum_value_eur avg_wage													
+	.+	+	+	+	+										
club	Real Madrid	1,050,650,000 118,934	24	38	[RW, CF,	RB,	LW, GK,	CAM,	RM, CB	, ST,	CDM,	LB,	CM]		
club	Manchester City	1,281,910,000 101,385	25	39	[RW, CF,	RB,	LW, GK,	RWB,	CAM, R	4, CB,	ST,	CDM,	LM,	LB, CM	1
club	FC Barcelona	1,131,875,000 91,571	24	49	[RW, CF,	RB,	LW, GK,	RWB,	CAM, C	3, RM,	ST,	CDM,	LM,	LB, CM]
league	La Liga	7,612,290,000 21,643	25	1143	[RW, CF,	RB,	LW, GK,	LWB,	RWB, C	₽M, CE	, RM	, ST,	CDM,	LM, L	B, CM
league	Premier League	11,049,435,000 21,002	24	2295	[[RW, CF,	RB,	LW, GK,	LWB,	RWB, C	M, RA	I, CB	ST,	CDM,	LM, L	B, CM

league	Serie A	7,729,900,000	19,340	26	1544	[RW, CM, CF, RB, LW, GK, LWB, RWB, CAM, RM, ST, CDM, LM, LB, CB
Inationalit	vlEgypt	195,570,000	17,681	26	47	[RW, CM, CF, RB, LW, RWB, CAM, RM, ST, CDM, LM, LB, CB]
nationalit	y Dominican Republi	10,030,000	13,800	23	10	[RW, CM, RB, GK, CAM, RM, ST, LM, LB, CB]
Inationalit	v Brazil	17,025,990,000	113.337	127	12440	I [RW, CF, RB, LW, GK, LWB, RWB, CAM, CB, RM, ST, CDM, LM, LB, CM
	9 1	. , , , ,	, ,			1,,,,,,,,,

Alternatywne rozwiązanie (nie do końca spójne)

category	Iname	sum_value_eur	avg_wage_eur	avg_age	count_player	player_positions
club	Real Madrid	1,050,650,000	118,934	24	38	[RN, CF, RB, LN, GK, CAM, RM, CB, ST, CDM, LB, CM]
club	Manchester City	1,281,910,000	101,385	25	39	[RN, CF, RB, LW, GK, RNB, CAM, RM, CB, ST, CDM, LM, LB, CM]
club	FC Barcelona	1,131,875,000	91,571	24	49	[RW, CF, RB, LW, GK, RMB, CAM, CB, RM, ST, CDM, LM, LB, CM]
league	Premier League	10,481,385,000	24,217	24	1959	[RN, CF, RB, LW, GK, LNB, RNB, CAM, RM, CB, ST, CDM, LM, LB, CM]
league	La Liga	7,612,290,000	21,643	25	1143	[RN, CF, RB, LW, GK, LNB, RNB, CAM, CB, RM, ST, CDM, LM, LB, CM]
league	Serie A	7,729,900,000	19,340	26	1544	[RN, CM, CF, RB, LN, GK, LNB, RNB, CAM, RM, ST, CDM, LM, LB, CB]
nationalit	y United Arab Emirate	s 10,620,000	20,250	24	2	[RB, CAM]
nationalit	ty Egypt	191,520,000	18,433	26	45	[RW, CM, CF, RB, LW, RWB, RM, CAM, ST, CDM, LM, LB, CB]
nationalit	y Dominican Republic	10,030,000	13,800	23	10	[[RH, CH, RB, GK, CAM, RM, ST, LM, LB, CB]

Zestaw 7 - carstrucks-data*

+	+		+	
sta	te manufacture	dominant_col	or avg_prio	e cars colors
+	+		+	++
GA	bmu	black	20,043	399 [blue, white, green, grey, brown, red, silver, black, custom, null]
CA	cadillac	null	17,015	[569 [yellow, blue, white, green, grey, brown, red, silver, black, custom, null, purple]
l ok	chevrolet	null	20,244	387 [orange, blue, white, green, grey, brown, red, silver, black, custom, null]
N3	dodge	null	16,523	122 [yellow, orange, blue, white, green, grey, brown, red, silver, black, custom, null]
ME	ford	white	20,449	197 [yellow, blue, green, white, grey, brown, red, silver, black, custom, null]
IN	gnc	null	20,954	305 [orange, blue, white, green, grey, brown, red, silver, black, custom, null]
NC.	honda	null	15,057	550 [blue, white, green, grey, brown, red, silver, black, custom, null, purple]
NE	hyundai	null	13,719	83 [[orange, blue, white, green, grey, brown, red, silver, black, null]
WY	jeep	null	24,292	134 [orange, blue, white, green, grey, brown, red, silver, black, custom, null]
MI	kia	null	14,357	462 [yellow, orange, blue, white, green, grey, brown, red, silver, black, custom, null, purple]
ID	lexus	null	19,544	254 [yellow, blue, white, grey, brown, red, silver, black, custom, null]
OH	mercedes-ber	nz black	20,757	433 [blue, white, grey, brown, red, silver, black, custom, null, purple]
NA	mitsubishi	null	19,264	173 [yellow, orange, blue, white, green, grey, brown, red, silver, black, custom, null]
SC	nissan	null	14,762	237 [orange, blue, white, green, grey, brown, red, silver, black, custom, null]
WY	ram	null	27,859	90 [blue, white, green, red, silver, black, custom, null]
l ok	subaru	null	21,547	117 [blue, green, white, grey, red, silver, black, null, purple]
VA	toyota	null	18,878	437 [blue, white, green, grey, brown, red, silver, black, custom, null, purple]
OH	volkswagen	black	14,773	326 [yellow, orange, blue, white, green, grey, red, silver, black, custom, null, purple]
+	+	+	+	++

Zestaw 8 - flight-delays

year [mon	ciijscai	relinosr_deraye	d_airline avg_de	tay [per cen	Clarit	Tiles											
+	+	*		+	-+												
2015 1	IL	F9	19	46.71	[F9,	В6,	DL,	US,	AS,	UA,	MQ,	00,	EV,	vx,	NK,	AA,	WN]
2015 2	MA	00	26	57.55	[B6,	DL,	US,	AS,	UA,	00,	EV,	NK,	vx,	AA,	WN]		
2015 3	NY	F9	21	46.01	[F9,	DL,	В6,	HA,	US,	UA,	MQ,	00,	EV,	٧X,	NK,	AA,	WN]
2015 4	NY	F9	14	37.92	[F9,	В6,	DL,	на,	US,	UA,	MQ,	00,	EV,	٧x,	NK,	AA,	WN]
2015 5	TX	NK	18	44.71	[F9,	В6,	DL,	US,	AS,	UA,	MQ,	00,	EV,	νx,	NK,	AA,	WN]
2015 6	N3	NK	24	46.57	[F9,	В6,	DL,	US,	AS,	UA,	MQ,	AA,	VΧ,	NK,	EV,	WN]	
2015 7	N3	F9	21	42.73	[F9,	В6,	DL,	AS,	UA,	MQ,	AA,	vx,	NK,	EV,	WN]		
2015 8	N3	F9	16	38.01	[F9,	В6,	DL,	AS,	UA,	MQ,	AA,	vx,	NK,	EV,	WN]		
2015 9	NY	HA	11	33.10	[F9,	В6,	DL,	на,	AS,	UA,	MQ,	00,	AA,	٧x,	EV,	NK,	WN]
2015 11	N3	AA	15	34.94	[F9,	DL,	В6,	AS,	UA,	MQ,	EV,	vx,	NK,	AA,	WN]		
2015 12	IN3	I B 6	119	139.64	IfF9.	В6,	DL.	AS.	UA.	мо.	AA.	vx.	NK.	EV.	WN1		

Zestaw 9 – discogs*

+	-+	+	-+	-+	+
decad	e label_name	artists_coun	t releases_count	t genres	best_year
+	-+	+	-+	+	+
1960	Columbia	3205	18733	[[Rock, Classical, Reggae, Jazz, Latin, Electronic, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1966
1960	RCA Victor	2189	16237	[Rock, Classical, Reggae, Latin, Jazz, Electronic, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul	1967
1960	Decca	1836	14421	[[Rock, Classical, Reggae, Jazz, Latin, Electronic, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1967
1970	Polydor	5665	43613	[[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1979
1970	CBS	4476	40907	[Rock, Classical, Reggae, Latin, Jazz, Electronic, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1979
1970	Columbia	4978	35782	[Rock, Classical, Reggae, Jazz, Latin, Electronic, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1972
1980	CBS	4574	48777	[Rock, Classical, Reggae, Latin, Jazz, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1989
1980	Polydor	4397	36366	[[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1980
1980	EMI	4400	34560	[Rock, Classical, Reggae, Latin, Jazz, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1988
1990	Not On Labe	1 47470	67076	[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Non-Music, Pop, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1999
1990	Columbia	3635	48627	[[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1991
1990	EMI	4115	28084	[Rock, Classical, Reggae, Latin, Jazz, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	1990
2000	Not On Labe	1 8803	10498	[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Brass & Military, Blues, Stage & Screen, Funk / Soul]	2000
2000	Columbia	649	3521	[Rock, Classical, Reggae, Jazz, Latin, Electronic, Hip Hop, Folk, World, & Country, Pop, Non-Music, Children's, Blues, Stage & Screen, Funk / Soul]	2000
2000	Epic	446	2735	[Pop, Rock, Reggae, Non-Music, Latin, Jazz, Blues, Electronic, Hip Hop, Stage & Screen, Folk, World, & Country, Funk / Soul]	2000
+	-+	+	.+	.++	+

Zestaw 10 - google-playstore-apps*

year developer_name				es sum_install	s categories
	4.15	16	773054	94000000	[Sports, Action, Adventure, Strategy, Casual, Role Playing, Puzzle, Arcade]
NULL An Phuc Vinh Company	3.75	6	22824	5000000	[[Sports, Arcade]
NULL Ajuntament de Barcelona	4.11	5	200453	5000000	[Puzzle]
2010 Refinery Productions	3.88	9	426566	45000000	[Sports, Casual, Puzzle]
2010 BraGIS BV	4.16	6	635932	33000000	[[Tools]
2010 Software Minder Limited	4.10	5	259236	27000000	[Puzzle, Word, Board]
2011 Nexsys Technologies	4.20	16	11694277	550000000	[[Libraries & Demo, Video Players & Editors]
2011 Fabio Anderi	4.20	5	206331	80000000	[[Educational]
2011 Interactive Health Solutions	4.27	12	400261	47000000	[[Lifestyle, Entertainment, Casual]
2012 Yash Wankhade	4.26	16	1575116	136000000	[[Music & Audio]
2012 . Iúri Batista Teles	4.19	5	6798457	130000000	[Action, Strategy, Casual]
2012 Dreamsoft Interactive Limited	1 4.27	6	715222	77000000	[Casual]
2013 . Iúri Batista Teles	4.31	8	31449241	341000000	[Card, Action, Strategy, Casual, Racing]
2013 Yama Lama Studios	4.41	16	1990938	140000000	[Music & Audio, Video Players & Editors, Books & Reference, Tools]
2013 Vector Mobile	4.22	10	966695	105800000	[Social, Lifestyle, Music & Audio, Entertainment, Education, Communication, Books & Reference, Medica
2014 Virgi Ing	4.41	16	722580	500000000	[Personalization, Photography, Productivity, Weather, Tools]
2014 Studio 1467	4.19	15	245952	346250000	[Lifestyle, Productivity, Education, Libraries & Demo, Weather, Communication, Books & Reference, Too
2014 AMANOTES PTE LTD	14.09	114	1924245	218100000	[Personalization, Lifestyle, Photography, Health & Fitness, Tools]
2015 IconLogin	4.34	9	5190625	350000000	[Photography]
2015 Studio 1467	4.13	24	179853	250800000	[Lifestyle, Social, Photography, Productivity, Entertainment, Video Players & Editors, Tools]
2015 Pak Appz	4.35	20	4829683	224000000	[Sports, Card, Action, Puzzle, Arcade, Word, Board]
2016 Realtorpreneur Ltd	4.61	6	1161021	2100700000	[Lifestyle, Productivity, Video Players & Editors, Tools]
2016 Fabio Anderi	14.27	110	799358	225000000	[[Educational]
2016 Freon Devs.	3.83	7	235476	200000000	[Photography, Entertainment, Tools]
2017 Freon Devs.	13.67	17	1112792	232000000	[Photography, Tools]
2017 Fabio Anderi	14.30	117	1568208	184000000	[[Educational]
	14.29	120	11618762	130300000	[[Sports, Strategy, Role Playing, Puzzle, Arcade, Board]
2018 Charli	14.06	116	18388170	1457500000	[Simulation, Action, Arcade]
2018 Mshamba Apps	4.16	İs	5175889	350000000	[[Action, Arcade]
	14.27	122	1709479	224000000	[[Educational]
2019 NDM	4.27	113	15525387	455000000	[[Simulation, Action, Casual, Puzzle]
	4.15	17	4126961	361100000	[[Sports, Puzzle, Arcade, Racing]
2019 Rovalturtle	4.16	118	1604351	306100000	[Simulation, Action, Casual, Role Playing, Puzzle, Arcade, Racing]
	14.05	111	1168102	386988888	[Simulation, Action, Casual, Arcade]
	14.56	17	1030282	280000000	[[Action, Casual, Arcade]
	13.87	113	11122246	266500000	[Sports, Simulation, Role Playing, Puzzle]
	4.24	15	188528	18000000	[[Educational]
	13.98	17	161757	17200000	[Sports, Action, Puzzle, Arcade]
	4.78	18	40981	650000	[[Simulation]

Last modified: Saturday, 9 November 2024, 18:37

Copyright @ 2024 Merito