گزارشکار پروژه داده‌کاوی

فاز ۱ و ۲

علی بدیعی

فایل‌های پروژه:

- **0.sqlMain.sql**

  فایل کوئری اصلی که با آن از سایت data.stackexchange.com دیتا ها را خارج کردیم.

- **1.QueryResults.csv**

  فایلی خروجی سایت با استفاده از کوئری ایجاد شده.

- **2.PythonResults.csv**

  فایل خروجی از پایتون که دیتاها را پروسس کرده و ستون‌هایی که در کوئری موجود نبود را اضافه کرد.

- **2.5.keeper.pickle**

  فایل خام لیبل زده‌شده بر اساس آیدی‌های هر سوال.

- **3.LabeledResults.csv**

  فایل خروجی جدول دیتاها که لیبل زده‌شده و آماده‌ی استفاده است.

- **4.PlotResults_1.png**
- **4.PlotResults_2.png**
- **4.PlotResults_3.png**

  خروجی‌های نمودار به صورت عکس.

- **main.py**

  فایل پایتون اصلی برای ۲ فاز اول پروژه.

نحوه‌ی انجام کار:

طبق موارد نوشته شده در متن پروژه عمل شد.

همچنین از فایل مقاله‌ی A Study on Classifying Stack Overflow Questions based on Difficulty by Utilizing Contextual Features برای موارد اضافی استفاده شد.

خروجی‌ها:

۱. خروجی کوئری اول

QuestionId	QuestionTi	QuestionBody	QuestionTags	QuestionBo	UserReputa	UserGoldBa	UserSilverB	UserBronze	QuestionVi	QuestionFa	UserUpVot	QuestionAn	QuestionSc	QuestionCreation	FirstAnswerCreation	AcceptedAnswerC	FirstAnswe	AcceptedAn
50093144	MySQL 8.0	I can't make a simple	<mysql><nod	2789	7655	3	10	14	1015467	0	7	35	755	18/04/30 2:04	18/05/02 9:58	18/05/02 9:58	2	2
49194719	Authentica	I am connecting	<mysql><mys	355	17755	9	53	82	1043189	0	715	38	671	18/03/09 13:19	18/03/12 5:08	18/03/12 5:08	3	3
50379839	Connectior	I try to connect	<java><mysql	2204	4667	3	9	10	670288	0	17	25	450	18/05/16 21:02	18/05/20 19:57	18/05/20 19:57	4	4
65456814	Docker (Ap	I'm running the	<mysql><doc	903	5470	4	31	52	405092	0	34	24	445	20/12/26 13:20	21/01/06 9:04	21/01/06 9:04	11	11
50177216	How to gra	Tried	<mysql><mys	753	2820	3	8	8	720528	0	7	24	282	18/05/04 14:23	18/05/06 8:17		2	
50690076	phpMyAdn	I have a WordPress	<php><mysql	620	5235	7	21	35	474164	0	9	22	272	18/06/04 23:15	18/06/07 16:58	18/06/07 16:58	3	3
52364415	PHP with M	I'm running MySQL	<php><mysql	542	2311	2	12	16	420053	0	1	8	228	18/09/17 9:17	18/09/17 9:18		0	
55038942	FATAL: pas	I recently installed	<postgresql><	306	1849	2	8	9	589784	0	0	18	183	19/03/07 8:12	19/03/07 8:41	19/03/07 8:41	0	0
50557234	Authentica	I am trying to	<python><my	768	4649	13	43	75	307207	0	215	29	172	18/05/27 22:53	18/05/27 23:05	18/05/27 23:05	0	0
50026939	php mysqli	I am using php	<php><mysql	1401	2568	2	16	20	310409	0	91	18	162	18/04/25 16:15	18/04/25 16:54	18/04/25 16:54	0	0
61523447	Skipping ac	Good afternoon,	<linux><postg	672	1601	2	6	3	108093	0	0	4	160	20/04/30 12:31	20/05/01 20:08		1	
55300370	PostgreSQL	To have an integer	<postgresql>	614	36202	11	45	69	79782	0	1709	1	150	19/03/22 13:10	19/03/22 13:31	19/03/22 13:31	0	0
57879150	How can I s	I want to run this	<node.js><sql	9671	12524	73	249	449	296271	0	463	28	144	19/09/10 22:59	20/01/23 16:12		135	
48218065	Objects cre	I can't find the	<python><my	1064	1293	2	7	5	143516	0	0	11	128	18/01/12 0:52	18/01/12 1:17	18/01/12 1:17	0	0
58461178	How to fix	Iâ€™m trying to	<postgresql>	613	1393	2	7	8	116287	0	5	15	126	19/10/19 6:39	19/10/19 7:19		0	
57665645	Server retu	My Intelij IDE wont	<mysql><sql>	277	1240	2	6	7	88473	0	54	5	113	19/08/26 22:29	19/08/27 2:27	19/08/27 2:27	1	1
54513450	A strange c	<ul>	<sql><sql-ser	1122	839	1	6	7	8047	0	1	3	105	19/02/04 9:43	19/02/04 11:02		0	
60864367	#1030 - Go	Created a new	<mysql><php	355	1003	2	8	7	156659	0	4	6	100	20/03/26 9:36	20/04/30 8:52	20/05/17 8:56	35	52
52032739	Loading cla	This is the Warning	<mysql><jdbc	335	1021	2	7	7	278424	0	0	26	100	18/08/27 4:52	18/09/19 12:23		23	
59993844	ERROR: Loa	I don't understand	<mysql><clier	2344	1121	1	6	4	257829	0	0	15	99	20/01/30 20:15	20/02/02 15:45		3	
51335298	Concepts o	Can anyone explain	<python><da	1052	1009	2	10	11	34417	0	0	2	97	18/07/14 4:44	20/01/26 17:15		561	
52423595	mysqldum	I want to dump my	<mysql>	631	14737	6	67	65	61214	0	64	3	93	18/09/20 11:00	18/09/20 11:00	18/09/20 11:00	0	0
50691977	How to res	I have actually lost	<mysql><pass	1294	1479	1	12	16	155526	0	42	8	93	18/06/05 4:20	18/07/05 8:29	18/09/30 16:29	30	117
48406228	Room - Sel	Is it possible to use	<android><sq	1458	3025	2	15	22	48798	0	29	3	89	18/01/23 16:14	18/01/23 16:14	18/01/23 16:14	0	0
54340470	TRIM is not	For the following	<sql><sql-ser	236	923	1	7	10	118870	0	3	3	89	19/01/24 6:21	19/01/24 6:26	19/01/24 6:26	0	0
49948350	phpMyAdn	UPDATE</st	<mysql><php	865	1690	1	14	29	195495	0	943	18	84	18/04/20 19:15	18/04/20 21:46	18/05/20 16:59	0	30
51442639	Why is 199	I have some clients	<sql-server><	3784	1493	0	14	26	6128	0	18	2	83	18/07/20 12:32	18/07/20 13:05	18/07/20 13:05	0	0

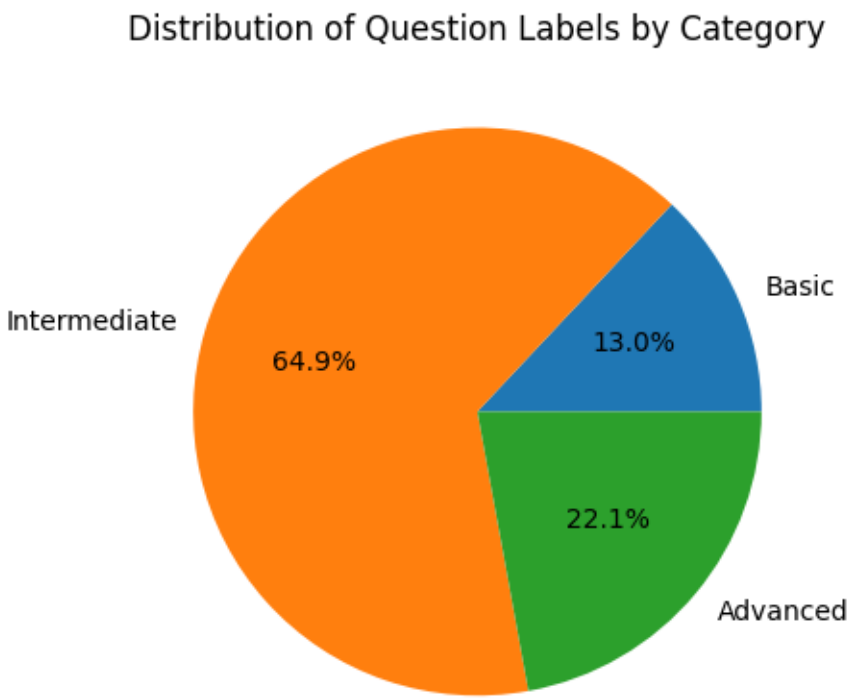
۲. خروجی فاز ۱ پایتون

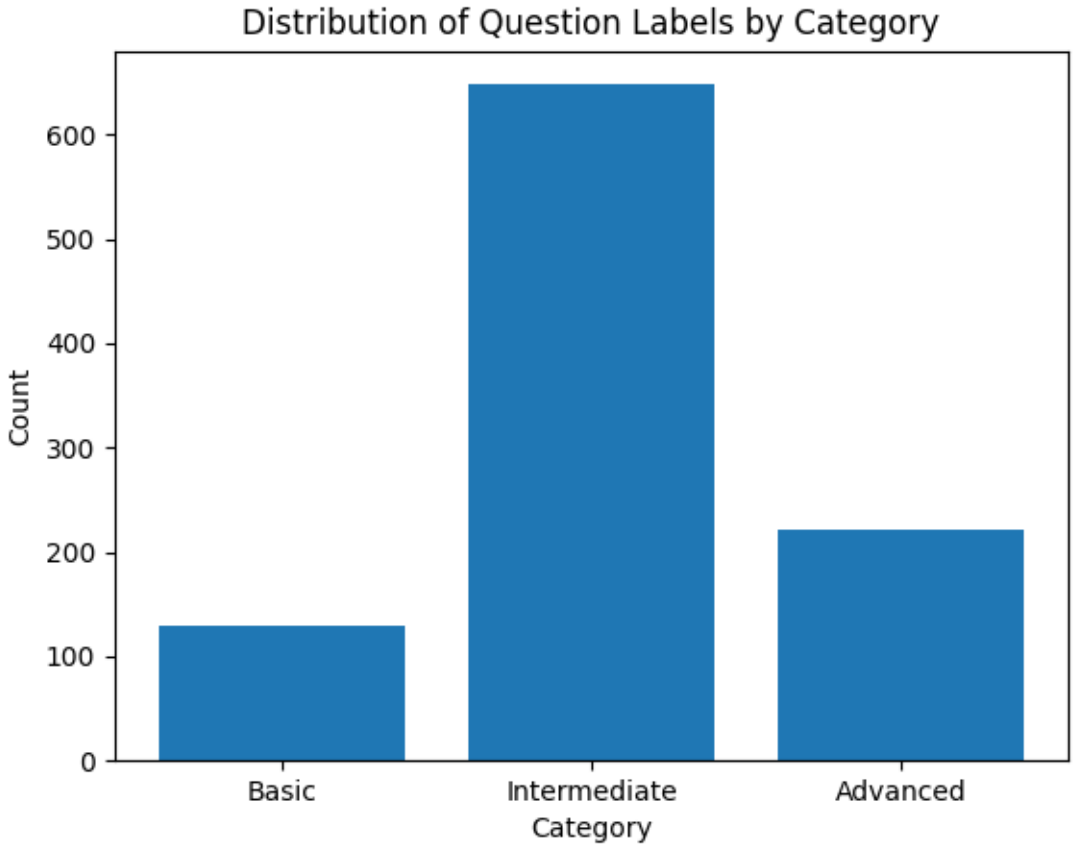
QuestionId	QuestionTi	QuestionBody	QuestionTa	QuestionBo	URLImageC	LOC	UserReputa	UserGoldBa	UserSilverB	UserBronze	QuestionAn	QuestionVi	QuestionFa	UserUpVot	QuestionAn	QuestionSc	QuestionCreation	FirstAnswe	AcceptedA	FirstAnswe	AcceptedAnsw
50093144	MySQL 8.0	I can't make a	<mysql><n	2492	4	51	7655	3	10	14	64	1015467	0	7	35	755	18/04/30 2:04	########	########	2	2
49194719	Authentica	I am connecting	<mysql><m	297	0	0	17755	9	53	82	61	1043189	0	715	38	671	18/03/09 13:19	########	########	3	3
50379839	Connectior	I try to connect	<java><my	2115	0	37	4667	3	9	10	73	670288	0	17	25	450	18/05/16 21:02	########	########	4	4
65456814	Docker (Ap	I'm running the	<mysql><d	730	1	16	5470	4	31	52	54	405092	0	34	24	445	20/12/26 13:20	########	########	11	11
50177216	How to gra	Tried	<mysql><m	584	0	2	2820	3	8	8	53	720528	0	7	24	282	18/05/04 14:23	########		2	
50690076	phpMyAdn	I have a	<php><my	539	0	2	5235	7	21	35	63	474164	0	9	22	272	18/06/04 23:15	########	########	3	3
52364415	PHP with M	I'm running	<php><my	441	0	0	2311	2	12	16	67	420053	0	1	8	228	18/09/17 9:17	########		0	
55038942	FATAL: pas	I recently	<postgresq	292	0	0	1849	2	8	9	71	589784	0	0	18	183	19/03/07 8:12	########	########	0	0
50557234	Authentica	I am trying to	<python><m	499	1	8	4649	13	43	75	42	307207	0	215	29	172	18/05/27 22:53	########	########	0	0
50026939	php mysqli	I am using php	<php><my	1069	1	22	2568	2	16	20	51	310409	0	91	18	162	18/04/25 16:15	########	########	0	0
61523447	Skipping ac	Good	<linux><po	548	3	3	1601	2	6	3	50	108093	0	0	4	160	20/04/30 12:31	########		1	
55300370	PostgreSQL	To have an	<postgresq	343	2	4	36202	11	45	69	71	79782	0	1709	1	150	19/03/22 13:10	########	########	0	0
57879150	How can I s	I want to run	<node.js><	9526	4	106	12524	73	249	449	65	296271	0	463	28	144	19/09/10 22:59	########		135	
48218065	Objects cre	I can't find the	<python><	981	0	15	1293	2	7	5	52	143516	0	0	11	128	18/01/12 0:52	########	########	0	0
58461178	How to fix	Iâ€™m trying to	<postgresq	466	0	2	1393	2	7	8	71	116287	0	5	15	126	19/10/19 6:39	########		0	
57665645	Server retu	My Intelij IDE	<mysql><s	262	0	0	1240	2	6	7	81	88473	0	54	5	113	19/08/26 22:29	########	########	1	1
54513450	A strange c		<sql><sql-s	727	1	13	839	1	6	7	61	8047	0	1	3	105	19/02/04 9:43	########		0	
60864367	#1030 - Go	Created a new	<mysql><p	237	1	0	1003	2	8	7	44	156659	0	4	6	100	20/03/26 9:36	########	########	35	52
52032739	Loading cla	This is the	<mysql><jc	303	0	4	1021	2	7	7	70	278424	0	0	26	100	18/08/27 4:52	########		23	
59993844	ERROR: Loa	I don't	<mysql><cl	2238	0	38	1121	1	6	4	38	257829	0	0	15	99	20/01/30 20:15	########		3	
51335298	Concepts o	Can anyone	<python><d	958	0	14	1009	2	10	11	68	34417	0	0	2	97	18/07/14 4:44	########		561	
52423595	mysqldum	I want to dump	<mysql>	510	0	1	14737	6	67	65	68	61214	0	64	3	93	18/09/20 11:00	########	########	0	0
50691977	How to res	I have actually	<mysql><p	999	2	11	1479	1	12	16	68	155526	0	42	8	93	18/06/05 4:20	########	########	30	117
48406228	Room - Sel	Is it possible to	<android><	1116	1	13	3025	2	15	22	76	48798	0	29	3	89	18/01/23 16:14	########	########	0	0
54340470	TRIM is not	For the	<sql><sql-s	159	0	4	923	1	7	10	57	118870	0	3	3	89	19/01/24 6:21	########	########	0	0
49948350	phpMyAdn	UPDATE	<mysql><p	773	0	3	1690	1	14	29	69	195495	0	943	18	84	18/04/20 19:15	########	########	0	30
51442639	Why is 199	I have some	<sql-server	3520	0	86	1493	0	14	26	45	6128	0	18	2	83	18/07/20 12:32	########	########	0	0

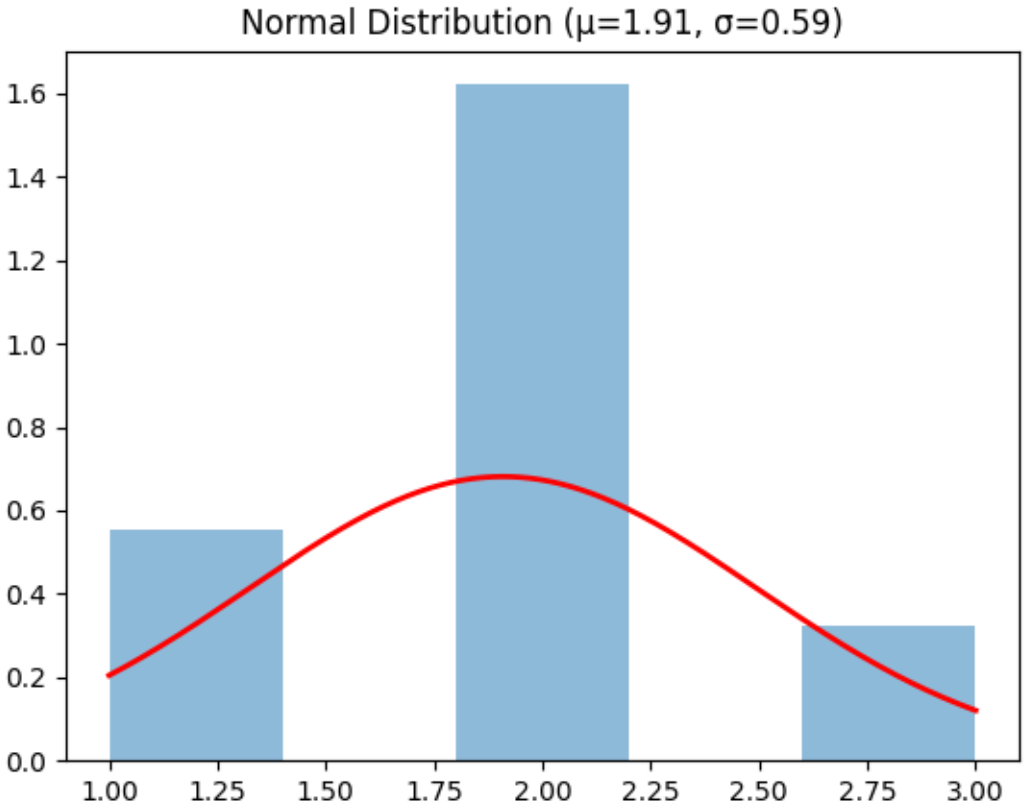
۳. خروجی لیبل‌زده شده

QuestionId	QuestionTit	QuestionBo	QuestionTa	QuestionBo	URLImageC	LOC	UserReputa	UserGoldBa	UserSilverB	UserBronze	QuestionAc	QuestionVie	QuestionFa	UserUpVot	QuestionAn	QuestionSc	QuestionCr	FirstAnswer	AcceptedAr	FirstAnswer	AcceptedAr	QuestionLa	QuestionLabelDefinition
50093144	MySQL 8.0 -	I can't	<mysql><nc	2492	4	51	7655	3	10	14	64	1015467	0	7	35	755	########	########	########	2	2	Advanced	32
49194719	Authenticat	I am	<mysql><m	297	0	0	17755	9	53	82	61	1043189	0	715	38	671	########	########	########	3	3	Advanced	32
50379839	Connection	I try to	<java><mys	2115	0	37	4667	3	9	10	73	670288	0	17	25	450	########	########	########	4	4	Intermediat	31
65456814	Docker (Apa	I'm	<mysql><dc	730	1	16	5470	4	31	52	54	405092	0	34	24	445	########	########	########	11	11	Advanced	37
50177216	How to gra	Tried	<mysql><m	584	0	2	2820	3	8	8	53	720528	0	7	24	282	########	########		2		Basic	9
50690076	phpMyAdm	I have a	<php><mys	539	0	2	5235	7	21	35	63	474164	0	9	22	272	########	########	########	3	3	Basic	14
52364415	PHP with M	I'm	<php><mys	441	0	0	2311	2	12	16	67	420053	0	1	8	228	########	########		0		Basic	3
55038942	FATAL: pass	I recently	<postgresql	292	0	0	1849	2	8	9	71	589784	0	0	18	183	########	########	########	0	0	Basic	13
50557234	Authenticat	I am trying	<python><r	499	1	8	4649	13	43	75	42	307207	0	215	29	172	########	########	########	0	0	Advanced	44
50026939	php mysqli_	I am using	<php><mys	1069	1	22	2568	2	16	20	51	310409	0	91	18	162	########	########	########	0	0	Advanced	44
61523447	Skipping acc	Good	<linux><pos	548	3	3	1601	2	6	3	50	108093	0	0	4	160	########	########		1		Basic	14
55300370	PostgreSQL	To have an	<postgresql	343	2	4	36202	11	45	69	71	79782	0	1709	1	150	########	########	########	0	0	Intermediat	19
57879150	How can I s	I want to	<node.js><s	9526	4	106	12524	73	249	449	65	296271	0	463	28	144	########	########		135		Intermediat	30
48218065	Objects crea	I can't find	<python><r	981	0	15	1293	2	7	5	52	143516	0	0	11	128	########	########	########	0	0	Basic	6
58461178	How to fix "i	iâ€™m	<postgresql	466	0	2	1393	2	7	8	71	116287	0	5	15	126	########	########		0		Basic	9
57665645	Server retu	My Intelij	<mysql><sq	262	0	0	1240	2	6	7	81	88473	0	54	5	113	########	########	########	1	1	Basic	14
54513450	A strange o		<sql><sql-se	727	1	13	839	1	6	7	61	8047	0	1	3	105	########	########		0		Basic	2
60864367	#1030 - Got	Created a	<mysql><ph	237	1	0	1003	2	8	7	44	156659	0	4	6	100	########	########	########	35	52	Advanced	37
52032739	Loading cla	This is the	<mysql><jdl	303	0	4	1021	2	7	7	70	278424	0	0	26	100	########	########		23		Advanced	38
59993844	ERROR: Loa	I don't	<mysql><cli	2238	0	38	1121	1	6	4	38	257829	0	0	15	99	########	########		3		Basic	3
51335298	Concepts of	Can	<python><c	958	0	14	1009	2	10	11	68	34417	0	0	2	97	########	########		561		Intermediat	27
52423595	mysqldump	I want to	<mysql>	510	0	1	14737	6	67	65	68	61214	0	64	3	93	########	########	########	0	0	Advanced	32
50691977	How to rese	I have	<mysql><pa	999	2	11	1479	1	12	16	68	155526	0	42	8	93	########	########	########	30	117	Basic	9
48406228	Room - Sele	Is it	<android><(	1116	1	13	3025	2	15	22	76	48798	0	29	3	89	########	########	########	0	0	Basic	9
54340470	TRIM is not	For the	<sql><sql-se	159	0	4	923	1	7	10	57	118870	0	3	3	89	########	########	########	0	0	Basic	1
49948350	phpMyAdm	UPDATE	<mysql><ph	773	0	3	1690	1	14	29	69	195495	0	943	18	84	########	########	########	0	30	Basic	14
51442639	Why is 199.	I have	<sql-server>	3520	0	86	1493	0	14	26	45	6128	0	18	2	83	########	########	########	0	0	Basic	2
48427185	How to mal	I've been	<dataframe	642	2	0	41968	17	109	152	50	9088	0	3145	4	83	########	########	########	0	0	Intermediat	31
50645402	MySQL: Syn	I just	<mysql><sq	285	0	2	2164	4	24	42	81	101109	0	417	7	82	########	########	########	0	0	Basic	1
57265913	Error: TCP P	I am trying	<sql-server>	725	2	2	821	1	6	3	43	102076	0	0	21	82	########	########		0		Intermediat	15
56108144	Using S3 as	Due to	<mongodb>	540	0	0	2978	6	26	47	80	63603	0	1087	2	80	########	########	########	0	0	Intermediat	18
50746465	How do I ca	I am	<c#><entity	2470	0	53	2850	2	17	34	65	57172	0	50	8	80	########	########	########	1	51	Basic	9

۴. خروجی نمودارها




Distribution of Question Labels by Category
Count
0
100
200
300
400
500
600
Basic
Intermediate
Advanced
Category


Normal Distribution (μ=1.91, σ=0.59)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00