#### Q1

#### HTTP:

```
HTTP日志结构:
root
 |-- host: string (nullable = true)
 |-- id.orig_h: string (nullable = true)
 |-- id.orig p: long (nullable = true)
 |-- id.resp_h: string (nullable = true)
 |-- id.resp_p: long (nullable = true)
  -- method: string (nullable = true)
 |-- orig_filenames: array (nullable = true)
     |-- element: string (containsNull = true)
 |-- orig_fuids: array (nullable = true)
     |-- element: string (containsNull = true)
  -- orig_mime_types: array (nullable = true)
      |-- element: string (containsNull = true)
  -- origin: string (nullable = true)
  -- proxied: array (nullable = true)
      |-- element: string (containsNull = true)
  |-- referrer: string (nullable = true)
 |-- request_body_len: long (nullable = true)
 |-- resp_filenames: array (nullable = true)
     |-- element: string (containsNull = true)
  -- resp_fuids: array (nullable = true)
     |-- element: string (containsNull = true)
  -- resp_mime_types: array (nullable = true)
     |-- element: string (containsNull = true)
  -- response_body_len: long (nullable = true)
 |-- status_code: long (nullable = true)
 |-- status_msg: string (nullable = true)
  -- tags: array (nullable = true)
     |-- element: string (containsNull = true)
 |-- trans_depth: long (nullable = true)
 |-- ts: timestamp (nullable = true)
 |-- uid: string (nullable = true)
  -- uri: string (nullable = true)
 |-- user_agent: string (nullable = true)
 |-- username: string (nullable = true)
 |-- version: string (nullable = true)
DNS:
DNS日志结构:
 |-- AA: boolean (nullable = true)
 |-- RA: boolean (nullable = true)
 |-- RD: boolean (nullable = true)
 |-- TC: boolean (nullable = true)
 |-- TTLs: array (nullable = true)
     |-- element: double (containsNull = true)
 -- Z: long (nullable = true)
 |-- answers: array (nullable = true)
      |-- element: string (containsNull = true)
  -- id.orig_h: string (nullable = true)
  -- id.orig_p: long (nullable = true)
  -- id.resp_h: string (nullable = true)
 |-- id.resp_p: long (nullable = true)
 |-- proto: string (nullable = true)
 |-- qclass: long (nullable = true)
 |-- qclass_name: string (nullable = true)
 |-- qtype: long (nullable = true)
 |-- qtype_name: string (nullable = true)
  -- query: string (nullable = true)
 |-- rcode: long (nullable = true)
 |-- rcode_name: string (nullable = true)
 |-- rejected: boolean (nullable = true)
 |-- rtt: double (nullable = true)
 |-- trans_id: long (nullable = true)
 |-- ts: timestamp (nullable = true)
 |-- uid: string (nullable = true)
```

## 使用 spark SQL API

uri	++  access_count
/	++   9475
/admin/config.php	556
/main.php?logout=1	194
/top.php?stuff=15	191
/top.php	179
/main.php?stuff=1	172
/get_latest_id.php	159
/admin/config.php	138
/cacti/index.php	129
/en-US/api/messag	118
/index.php	105
/phpmyadmin/index	77
/cacti/	68
/phpmyadmin/	56
/favicon.ico	55
/admin	42
/scripts/index.php	40
/icons/	39
/en-US/api/search	39
/cgi-bin/index.php	37
+	++

only showing top 20 rows

#### 使用 spark DataFram

uri	count
	19475
//admin/config.php?type=tool&display=index&quietmode=1&info=stats&restrictmods=core/dashboard	556
/main.php?logout=1	194
/top.php?stuff=1583574484	191
/top.php	179
/main.php?stuff=1583574484	172
/get latest id.php	159
/admin/config.php?type=tool&display=index&quietmode=1&info=info&restrictmods=core/dashboard	138
/cacti/index.php	129
/en-US/api/messages/index	118
/index.php	105
/phpmyadmin/index.php	77
/cacti/	68
/phpmyadmin/	56
/favicon.ico	55
/admin	42
/scripts/index.php	40
/icons/	39
/en-US/api/search/jobs?s=1331892438.21&s=1331892438.24	39
/cgi-bin/index.php	37

only showing top 20 rows

#### 合并并计算验证

实现参考:

https://notebooks.databricks.com/notebooks/CME/Survival\_Analysis/index.html#Survival\_Analysis 1.html

首先从以下链接下载 csv 文件至服务器

 $\underline{raw.githubusercontent.com/IBM/telco-customer-churn-on-icp4d/master/data/Telco-Customer-Churn.csv}$ 

该 csv 数据文件来自 IBM,旨在模拟一个虚构的电信公司的用户数据。每一行数据都代表着一个电信公司的订阅用户的个人信息,包括但不仅限于各自人口统计、服务计划、媒体使用情况、订阅状态以及在网时长和是否已经流失的信息。

我们要对这组数据进行生存分析,最重要的内容即是在网时长(Tenure)和是否流失(Churn),从对这两组信息的分析,我们可以后续进行估计判断客户留存与在网时长的关系或某时间点内客户未流失的概率,进而对客户的市场需求和改良电信产品做出应对措施等。

我们会根据用户的订阅模式,将用户数据分为 bronze 和 silver 两部分,主要分析 silver 高级用户的留存率,便于指定针对化的措施。

#### 具体执行方案:

前期配置,创建spark、文件路径和表名等。然后从csv数据文件的表头获取schema。首先将全体数据都存在 bronze 中,进行预存储:

+									
customerID gender seniorCi ackup  deviceProtection  monthlyCharges totalCharges	tizen pa	techSupport		streamingTV	stream	ningMovies	contract paperle	essBilling	paymentMethod
+									
		(Section 1	2000		. In		and t		witt
7590-VHVEG Female	0.0	Yes	No	1.0	No   No p	hone service		VI	No  Electronic check
Yes   No   29.85   29.85   No		No		No		No Month-t	.o-montn	Yes	Electronic check
ISSTS-GNVDE  Male	0.01	Nol	Nol	34.0	Yesl	Nol	DSL		Yesl
Nol Yesl	0.01	Nol	110	Nol	100		ne year	Nol	Mailed check
56.95  1889.5  No		,				,	,1	,	
3668-QPYBK  Male	0.0	No	No	2.0	Yes	No	DSL		Yes
Yes  No		No		No		No   Month-t	o-month	Yes	Mailed check
53.85 108.15 Yes									
7795-CFOCW  Male	0.0	No	No	45.0	No No p	ohone service			Yes
No  Yes		Yes		No		No On	ne year	No Ban	k transfer (au
42.3  1840.75  No	1	27.3	20 F	1		W 1			500 C
9237-HQITU Female	0.0	No	No	2.0	Yes	No  No Month-to	Fiber optic	Yesl	No  Electronic check
No  No  70.7  151.65  Yes		No		No		NO   MON CHI - CO	o-month	res	Electronic check
9305-CDSKC Female	0.01	Nol	Nol	8.0	Yes	Yes	Fiber optic		Nol
No Yes	0.01	Nol		Yes		Yes   Month-to		Yes	Electronic check
99.65   820.5   Yes							International Control	000001	
1452-KIOVK  Male	0.0	No	Yes	22.0	Yes	Yes	Fiber optic		No
Yes   No		No		Yes		No Month-t	o-month	Yes Cr	edit card (auto
89.1  1949.4  No									
6713-OKOMC Female	0.0	No	No	28	No No p	hone service	DSL	22.1	Yes
No   No		No		No		No   Month-to	-month	No	Mailed check

接着对于 silver 用户,筛选所有拥有月度订阅(month-to-month)的在网用户:

| customerID|gender|seniorCitizen|partner|dependents|tenure|phoneService| multipleLines|internetService|onlineSecurity|onlineBackup|device | Protection|techSupport|streamingTV|streamingMovies| contract|paperlessBilling| paymentMethod|monthlyCharges|totalCharges|

	+	+								+-		+	+
7590-VH	VEG Female	0.0	Yes	No	1.0	No I	No ph	none service		DSL		No	Yes
No	No	No	No Me	onth-to-m	nonth	Ye	es	Electronic	check		29.85	29.85	
3668-QP	YBK  Male	0.0	No	No	2.0	Yes		No		DSL		Yes	Yes
No	No	No	No M	onth-to-m	nonth	Ye	es	Mailed	check		53.85	108.15	
9237-HQ	]ITU Female	0.0										No	
No	No	No	No M	onth-to-m	nonth	Ye	es	Electronic	check		70.7	151.65	
9305-CD	SKC Female	0.0		No				Yes					No
Yes	No	Yes	Yes	Month-to-	month	,	Yes	Electronic	check		99.65	820.5	
1452-KI	OVK   Male	0.0						Yes					Yes
No	No	Yes	No M	onth-to-m	nonth	Ye	es   Cr	redit card (au	to		89.1	1949.4	
6713-OK	OMC Female	0.0	No	No	10.0	No I	No ph	none service		DSL		Yes	No
No	No	No						Mailed					
	OKP Female		Yes	No	28.0	Yes		Yes	Fiber	optic		No	No
Yes	Yes	Yes	Yes	Month-to-	month	,	Yes	Electronic	check		104.8	3046.05	
9763-GR	SKD  Male	0.0	Yes	Yes	13.0	Yes		No		DSL		Yes	No
No	No	No	No M	onth-to-m	nonth	Ye	es	Mailed	check		49.95	587.45	
0280-XJ	GEX   Male	0.0	No	No	49.0	Yes		Yes	Fiber	optic		No	Yes
Yes	No	Yes	Yes	Month-to-	-month	,	Yes   E	Bank transfer	(au		103.7	5036.3	
5129-JL	PIS   Male	0.0	No	No	25.0	Yes		No	Fiber	optic		Yes	No
Yes	Yes	Yes											
4190-MF	LUW Female	0.0	Yes	Yes	10.0	Yes		No		DSL		No	No
Yes	Yes	No	No   I	Month-to-	-month		No	Credit card (a	uto		55.2	528.35	
4183-MY	FRB Female	0.0	No	No	21.0	Yes		No	Fiber	optic		No	Yes
Yes	No	No	Yes	Month-to-	-month	,	Yes	Electronic	check		90.05	1862.9	
8779-QR	RDMV  Male	1.0	No	No	1.0	No I	No ph	none service		DSL		No	No
		No										39.65	
	RPFA  Male	0.0						No				Yes	Yes
Nol	Yes	No						redit card (au				2970.3	

对旧数据进行删除并创建新数据库,分别将 bronze 和 silver 用户数据写入。通过 spark SQL API 显示用户数据:

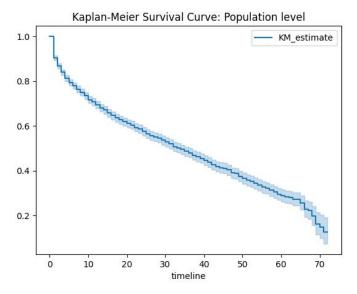
	+		+			+				-+		+	
ustomerID gender se							rvice  mu stream:				onlines essBilling	Security	onlin
<pre>(up  deviceProtec nthlyCharges totalC</pre>			echSupport		stream	urug i v l	stream.	Tulkuoviezi	contrac	clbaberi	ezzpititili	si paym	entMetho
			+	+	+		+	+-		+			
+	+		+			+				-+		+	
	+-	+											
590-VHVEG Female		0.0	Yes	No	1.0		No No ph	none service		DSL		No	
	No .		No			No		No   Month-	to-month		Yes	Electroni	c check
.85 29.85	No	0.01	no.	es i	24.01		Yes	w. I		DSLI		Yes	
575-GNVDE  Male    Ye	es	0.01	No  No	No	34.0	No	Yes	Nol C	ne year	DSL	No		check
.95 1889.5	Nol		NOT			NOT		NO C	ne year		NOT	Mailed	CHECK
668-QPYBK  Male		0.0	Nol	Nol	2.0		Yes	Nol		DSL		Yes	
	No		No			No		No   Month-	to-month		Yes		d check
	Yes												
795-CFOCW   Male		0.0	No	No	45.0		No No ph	none service		DSL		Yes	
Ye Ye			Yes			No		No O	ne year		No Ba	ank transfer	(au
	No				7				- 4.				
237-HQITU Female	lol	0.0	No  No	No	2.0	Nol	Yes	No  No Month-t		optic	Yes	No  Electronic	aha ak l
,	(es		NO			140		No month-t	o-month		Yes	stectrouic	cneck
305-CDSKC Female		0.0	No	Nol	8.01		Yes	Yes	Fiber	optic		Nol	
	es	- 101	No	1101		/es		Yes Month-t			Yes	Electronic	checkl
.65  820.5											0.00000		
L452-KIOVK  Male		0.0	No	Yes	22.0		Yes	Yes	Fiber	optic		No	
	No		No			Yes		No   Month-	to-month		Yes 0	Credit card (	auto
	No												
713-OKOMC Female		0.0	No	No	10.0		No No ph	none service		DSL	57.1	Yes	
N .75  301.9	lo  Nol		No			No		No   Month-t	o-month		No	Mailed	check
		+-	+	+			.+					+	+
	 + eniorCit	+- izen p	+ artner depe	ndents t	enure	phones	-+ Service	multipleLines	internet	Service	onlineSec	urity online	+ Backup d
ustomerID gender so	 eniorCit rt strea	+- izen p mingTV	+ artner depe   streamingM	ndents t	enure	phones	-+ Service  t paperless	multipleLines Billing	internet  payment	:Service  :Service  :Method n	onlineSec	urity online	Backup d
ustomerID gender s@	 eniorCit rt strea	izen p mingTV	+ artner depe   streamingM +	ndents t	enure c	phones	Service  t paperless	multipleLines Billing	internet  payment	Service	onlineSec	urity online	Backup darges ch
ustomerID gender si	+ eniorCit rt strea	+- izen p mingTV	+ artner depe   streamingM + +	ndents t  ovies   +  No	enure c	phones	Service    paperless   No No	multipleLines Billing	internet  payment  +	:Service  :Service  :Method n	onlineSec	urity online  rges totalCh	Backup d arges ch + Yes
ustomerID gender se otection techSuppor	 eniorCit rt strea	izen p mingTV +- 0.0	+	ndents t  ovies   +  No   nth-to-m	enure c  1.0	phones	Service    paperless   No No   Yes	multipleLines Billing  	internet  payment  t        c check	Service Service Method n	onlineSec	urity online  rges totalCh+ No  29.85	Backup darges chi
ustomerID gender sootection techSuppor	eniorCit rt strea	izen p mingTV	+	ndents t  ovies  +   No   nth-to-n	enure c 1.0 nonth	phones	Service  t paperless No No Yes	multipleLines Billing  phone service Electroni	internet  payment  +      c check	Service	onlineSectionthlyCha	urity online rges totalCh + No  29.85  Yes	Backup de arges che +- Yes  0  Yes
ustomerID gender so otection techSuppor 	+ eniorCit rt strea	izen p mingTV +- 0.0	+	ndents t  ovies   +  No   nth-to-m	enure c 1.0 nonth	phones	Service    paperless   No No   Yes	multipleLines Billing   phone service Electroni No Maile	internet  payment  +   c check   d check	zService  tMethod n DSL	onlineSec nonthlyCha 29.85	urity online  rges totalCh+ No  29.85	Backup de arges che +- Yes  0  Yes  1
ustomerID gender sc otection techSuppor	eniorCit rt strea	izen p mingTV +- 0.0	artner depe   streamingM 	ndents t lovies  + No  nth-to-m	1.0 nonth  2.0 nonth  2.0	phones	Service  t paperless No No Yes  Yes	multipleLines Billing   phone service Electroni No Maile	internet  payment     c check   d check	Service Service Method n	onlineSec nonthlyCha 29.85	urity online rges totalCh + No  29.85  Yes  108.15	Backup de arges che +- Yes  0  Yes
ustomerID gender scotection techSuppor	eniorCit rt strea	izen p mingTV +- 0.0	artner depe   streamingM 	ndents t  ovies  +   No   nth-to-m   No   nth-to-m	1.0 100nth  2.0 100nth	phonesontract	Service  t paperless  No No Yes  Yes  Yes  Yes	multipleLines Billing   phone service Electroni No Maile	internet  payment  +	zService  tMethod n DSL	onlineSec nonthlyCha 29.85  53.85  70.7	urity online rges totalCh + No  29.85  Yes  108.15  No	Backup d arges ch +- Yes  0  Yes  1  No
ustomerID gender sv otection techSuppor 590-VHVEG Female    No  668-QPYBK  Male    No  237-HQITU Female    No  385-CDSKC Female  s  No	eniorCit rt strea	izen pmingTV 0.0  0.0  0.0  0.0	+	ndents t lovies  No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n	1.00 100nth  2.00 100nth  8.00 100nth  8.00	phonesontract	No No Yes  Yes  Yes  Yes  Yes  Yes	multipleLines Billing   phone service Electroni No Maile No Electroni Yes	internet  payment 	DSL DSL	onlineSec nonthlyCha 29.85  53.85  70.7  99.65	unity online rges totalch No  29.85  Yes  168.15  No  151.65  No  820.5	Backup darges ch
ustomerID gender scotection techSuppor	No  No  Yes	izen pmingTV+- 0.0  0.0  0.0	+	ndents t  lovies +- No  nth-to-n No  nth-to-n No  onth-to-yes	1.0 nonth  2.0 nonth  8.0 month  22.0	phonesontract	No No Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes	multipleLines Billing   phone service Electroni No Maile No Electroni Yes   Electroni	internet payment	DSL DSL	onlineSec nonthlyCha 29.85  53.85  70.7  99.65	unity online rges totalCh No  29.85  Yes  108.15  No  151.65  No  820.5	Backup danges chi
ustomerID gender so totection techSuppor  599-VHVEG Female    No  668-QPYBK  Male    No  357-HQITU Female    No  355-CDSKC Female  s  No  452-KOVK  Male	eniorCit rt strea  No  No	izen pmingTV+ 0.0  0.0  0.0  0.0  0.0	+	ndents t  lovies  No  nth-to-n No  nth-to-n No  nth-to-n Yes	1.00 100nth  2.00 100nth  8.00 100nth  8.00 100nth  22.00 100nth  100n	phones	No No Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes	multipleLines :Billing	internet	Service   Service   Method n DSL   DSL   Proptic	onlineSec nonthlyCha 29.85  53.85  70.7  99.65  89.1	urity online  rges totalCh  No  29.85  Yes  108.15  No  151.65  No  820.5  No  1949.4	### Description of the content of th
ustomerID gender scotection techSuppor	eniorCit rt strea No  No  No  Yes	izen pmingTV 0.0  0.0  0.0  0.0	+	ndents t  lovies +- No  nth-to-m No  nth-to-m No  nth-to-m Yes  nth-to-m No	1.00 1001th  2.00 1001th  8.00 1001th  8.00 1001th  10.00	phones	No No Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes	multipleLines Billing   phone service Electroni No Maile No Electroni Yes   Electron	internet payment	DSL DSL	onlineSec ionthlyCha 29.85  53.85  70.7  99.65  89.1	unity online  rges totalCh	Packup d arges ch
ustomerID gender s/ otection techSuppor	No  No  Yes	izen pmingTV+- 0.0  0.0  0.0  0.0  0.0  0.0	+		1.00 100nth  2.00 100nth  2.00 100nth  8.00 100nth  22.00 100nth  10.00 100nth  10.00	phones	Service	multipleLines Billing   phone service Electroni No Maile No Electroni Yes   Electroni Yes Credit card ( phone service Maile	internet payment	Service   Service   Method n DSL   DSL  er optic   er optic   DSL   DSL	onlineSec ionthlyCha 29.85  53.85  70.7  99.65  89.1  29.75	No  29.85  Yes  188.15  No  151.65  No  820.5  No  1949.4  Yes  301.9	Backup d arges ch+ Yes  0  Yes  1  No  1  Yes  0  No  0  No
ustomerIO   gender  so tection  TechSuppor  599-VHYEG  Female    No  668-QPYBK  Male    No  305-CDSKC  Female  s  No  452-KIOVK  Male    No  713-OKOMC  Female    No  892-POOKP  Female	eniorCit rt strea No  No  No  Yes  Yes	izen pmingTV+ 0.0  0.0  0.0  0.0  0.0	+	ndents tiovies  No estable   No nth-to-mestable   No	1.00 100nth  2.00 100nth  2.00 100nth  8.00 100nth  10.00 100nth  10.00 100nth  28.00	phonesontract	No No No Yes	multipleLines Billing   phone service Electroni No Electroni Yes Credit card ( phone service Maile	internet payment 	DSL	onlineSec ionthlyCha 29.85  53.85  70.7  99.65  89.1  29.75	unity online rges totalCh 	Backup danges chi
ustomerID gender si otection techSuppor 5990-VHVEG Female    No  668-QPYBK  Male    No  237-HQITU Female    S  305-CDSKC Female    No  3452-KDVK  Male    No  713-0KOMC Female    No  892-POOKP Female    S  892-POOKP Female	eniorCit rt strea No  No  No  Yes	izen pmingTV 0.0  0.0  0.0  0.0  0.0  0.0  0	+	ndents t  ovies  No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  onth-to-n No	1.00 100nth  2.00 100nth  2.00 100nth  2.00 100nth  10.00 100nth  10.00 100nth  10.00 100nth  10.00 100nth  28.00	phonesontract	Holinois Andrews Andre	multipleLines Billing   phone service Electroni No Maile Electroni Yes   Electron Credit card ( phone service Maile	internet payment    c check    Fibe c check    Fibe ic check    Fibe auto      d check    Fibe ic check	DSL	onlineSeconorthlyCha- 29.85  53.85  70.7  99.65  89.1  29.75	No  29.85  Yes  188.15  No  151.65  No  820.5  No  1949.4  Yes  301.9	Backup d arges ch+ Yes  0  Yes  1  No  1  Yes  0  No  0  No
ustomerID gender sc otection techSuppor 5590-VHYEG Female    No  688-QPYBK   Male    No  237-HQITU Female    No  365-CDSKC Female    No  452-KIOVK  Male    No  713-OKOMC Female    No  892-POOKP Female    S  892-POOKP Female    S  892-MIDMERIES    Wes  810-MIDMERIES    Wes  810-MIDMERIE	No  No  No  Yes  No  Yes	izen pmingTV+- 0.0  0.0  0.0  0.0  0.0  0.0	artner   depe   streaming M	ndents t  ovies  No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  nth-to-n No  onth-to-n No	1.00 conth  2.00 conth  8.00 conth  10.00 conth  10.00 conth  10.00 conth  128.00 conth  13.00 c	phonesontract	No No No Yes	multipleLines Billing   phone service Electroni No Maile No Electroni Yes Credit card (phone service Maile Yes Electron	internet payment    c check    Fibe c check    Fibe ic check    Fibe auto      d check    Fibe ic check	DSL   Proptic   Proptic	onlineSeconorthlyCha- 29.85  53.85  70.7  99.65  89.1  29.75	unity online rges totalch	Backup d anges ch
ustomerID gender s otection techSuppor 590-VHVEG Female    No  668-QPYBK  Male    No  237-HQITU Female    No  305-CDSKC Female    No  452-KIOVK  Male    No  713-OKONC Female    No  773-OKONC Female    Yes  763-GRSKD  Male    No	eniorCit rt strea No  No  No  Yes  Yes	izen pmingTV 0.0  0.0  0.0  0.0  0.0  0.0  0	artner   depe   streaming M	ndents t  lovies  No  nth-to-m No  nth-to-m No  nth-to-m No  nth-to-m No  lonth-to- Yes  nth-to-m No  nth-to-m	1.00 conth  2.00 conth  8.00 conth  10.00 conth  10.00 conth  10.00 conth  128.00 conth  13.00 c	phones	No No Yes Yes Yes Yes Yes No No No No Yes	multipleLines Billing   phone service Electroni No Maile No Electroni Yes Credit card (phone service Maile Yes Electron	internet payment     c check       d check     Fibe     c check     Fibe     ic check     Fibe     auto       d check     Fibe     auto       d check     Fibe     d check     Fibe	DSL   Proptic   Proptic	29.85  53.85  70.7  99.65  89.1  29.75  104.8	Nol 29.85  Yes  188.15  No  151.65  No  1949.4  Yes  301.9  No  3646.05  Yes	Backup d arges ch
ustomerID gender sc otection techSuppor 599-VHVEG Female    No  668-QPYBK  Male    No  395-CDSKC Female    No  395-CDSKC Female    No  452-KIOVK  Male    No  91-713-OKONC Female    No  82-POXP Female    S  No  82-POXP Female    Yes  763-GRSKD  Male    No  8280-XJGEX  Male    No	No  No  No  Yes  No  Yes	izen pmingTV	artner   depe   streamingM   ves   No   Mo   No   No   No   No   No   No	ndents t  lovies	1.00 conth  2.00 conth  2.00 conth  2.00 conth  2.00 conth  10.00 conth  10.00 conth  40.00 conth	phones	Hervice	multiplelines Billing   phone service Electroni No Maile Yes Credit card ( phone service Maile Yes   Electron Mo Maile And Mo	internet payment     c check      d check    Fibe     c check    Fibe     d check    Fibe     d check    Fibe     d check    Fibe     f check    Fibe     r (au)	DSL   Proptic   Proptic   DSL   Proptic   Proptic   DSL   DSL   DSL   Proptic   DSL	29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95	vity online rges totalch No  29.85  Yes  188.15  No  151.65  No  828.5  No  1949.4  Yes  301.9  No  587.45  No  5936.3	Backup darges ch
ustomerID gender so totection techSuppor 599-VHYEG Female    No  688-QPYBK Male    No  237-HQITU Female    No  452-KIOVK  Male    No  713-OKOWC Female    No  882-POOKP Female  s  No  882-ROOKP Female  s  No  882-ROOKP Male    No  883-ROOKP Male    No  883-ROOKP Male    No  883-ROOKP Male    No  883-ROOKP Male    No	Nol Nol Yes  Nol Yes		Yes   No   Mo   No   No   No   No   No   No	ndents tiovies  No  nth-to-n No	1.0 cenure c	phones	No No Yes	multiplelines Billing   phone service Electroni No Maile No Electroni Yes   Electron Moile Yes   Electron Moile Yes   Electron No Maile   Electron No Maile No Maile Yes	internet   payment	DSL   Proptic	29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95	ves   ves	Backup d anges ch
ustomerID gender si otection techSuppoi  599-VHVEG Female    No  668-QPYBK  Male    No  335-CDSKC Female    No  335-CSKC  Female    No  3452-KIOVK  Male    No  713-0KOMC Female    No  892-POOKP Female    No  892-POOKP Female    No  892-POOKP Male    No  113-31-MOMC SEMALE   No  113-31-MOMC SEMALE   Ves    No  113-31-MOMC SEMALE   No  113-31-PIS  Male    No  113-31-PIS  Male    S  Yes	Nol Yesl Nol Yesl		Yes  No Mo No  No Mo No  No Mo No  No Mo No  Yes M No  No Mo No  No Mo No  No Mo No  No Mo No  Yes M No  Yes M No Mo Yes  Yes M No Mo No  No Mo No Mo Yes Mo No Mo No Mo No Mo Yes Mo No Mo		1.0 donth  2.0 donth  2.0 donth  10.0 donth  10.0 donth  13.0 donth  13.0 donth  25.0 month 13.0 donth  15.0 donth	phones	No No No Yes Yes Yes No No No No No Yes Yes Yes Yes Yes Yes Yes Yes Yes No No No Yes	multipletines Billing   phone service Electroni No Maile No Electroni Yes Credit card (phone service Maile Yes Electron No Maile Yes Electron No Maile Yes Electron	internet payment	DSL   Proptic  Proptic  Proptic  Proptic  DSL   Proptic  Proptic  Proptic  Proptic  Proptic	29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7	Nol 1949.4  Yes  391.9  Nol 3946.95  Yes  397.45  Nol 1949.4  Yes  391.9  Nol 3946.95  Yes  2686.95	Backup darges ch-
ustomerID gender s otection techSuppor 590-VHVEG Female    No  668-QPYBK  Male    No  237-HQITU Female    No  305-CDSKC Female    No  452-KIOVK  Male    No  713-OKONC Female    No  892-POOKP Female  s  Yes  763-GRSKD  Male    No  280-XJGKZ  Male    No  129-JLPIS  Male  s  Yes    Yes	Nol Nol Nol Yes  Yes  Nol Yes		res   No   Mo   No   No   No   No   No   No	ndents t  lovies	1.00 nonth  2.00 nonth  8.00 month  10.00 nonth  10.00 nonth  10.00 nonth  10.00 nonth  10.00 nonth  25.00 month 13.00 nonth  10.00 non	phones	No No No Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes	multiplelines Billing   phone service Electroni No Maile No Electroni 'yes Credit card ( phone service Maile Yes Electroni No Maile Service Maile Service Maile Service No Maile Service No Maile No	internet payment	DSL	onlineSec tonthlyCha 29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7	vity online   regs totalch   29.85    Yes    188.15    No    151.65    No    1949.4    Yes    301.9    No    3046.05    Yes    S87.45    No    5087.63    Yes    S086.05	Backup darges ch
ustomerID gender sc otection techSuppor 5590-VHYEG Female    No  668-0PYBK  Male    No  335-CDSKC Female    No  335-CDSKC Female    No  713-OKOMC Female    No  982-POOKP Female    No  182-POOKP Female    No  882-POOKP Female    No  182-POOKP Female    No  182-TJGEX Male    No  1890-TJGEX Male    No  1990-HPLUW Female    Yes    1990-MFLUW Female    No	Nol Nol Yes  Nol Yes		Yes   No   Mo   No   No   No   No   No   No	ndents tiovies	1.00 conth  2.00 conth  2.00 conth  2.00 conth  2.00 conth  1.00 c	phones	ervice	multiplelines Billing   phone service Electroni No Maile No Electroni Yes Credit card (phone service Maile Yes: Electron No Maile Yes Electron No Maile Helectron No Credit card	internet payment	DSL	29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  105.5	Nol 29.85  Yes  188.15  No  820.5  No  1949.4  Yes  301.9  No  9836.3  Yes  0.528.35	Backup daarges ch
ustomerID gender s otection techSuppor 599-VHVEG Female    No  688-QPYBK  Male    No  237-HQITU Female    No  305-CDSKC Female    No  305-CDSKC Female    No  452-KIOVK  Male    No  892-POXP Female    Ves  763-GRSKD  Male    No  289-YDGKP Female    No  280-XJGEX  Male    129-JIPIS  Male    129-JIPIS  Male    S  No  3129-JIPIS  Male    Yes  199-MFLUW Female    S  Yes  198-MFLUW Female    S  Yes	Nol Nol Nol Yes  Nol Yes  Yes  Nol Yes		Tren'   depe     streamingM     Yes   No   Mo     Yes   M     No   Mo     Yes   M     No   Mo     No	numbers   tovies   No	1.00 conth  2.00 conth  2.00 conth  2.00 conth  3.00 conth  10.00 conth  13.00 conth  15.00 cont	phones	No No No Vess Yes Yes Yes Yes Yes Yes Yes Yes Yes	multiplelines Billing   phone service Electroni No Maile Flectroni Yes Credit card ( phone service Maile Yes Electron No Maile Flectroni No Electron No Maile Flectron No Maile No Credit card No No Maile Rock Rock Rock Rock Rock Rock Rock Rock	internet payment   c check    d check    Fibe c check    Fibe ic check    Fibe id check    Fibe id check    Fibe ic check    Fibe ic check    Fibe ic check    Fibe ic check    Fibe   d check    Fibe   d check    Fibe   Gauto   Fibe   Gauto   Fibe	DSL   Proptic	onlineSec tonthlyChar 29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7  105.5	ves  No  198.45  No  29.85  Yes  108.15  No  151.65  No  1949.4  Yes  301.9  No  3046.05  Yes  2686.05  No  528.35  No	Backup d anges ch
ustomerID gender se otection techSuppor	Nol Nol Nol Yes  Yes  Nol Yes	izen pingTV 0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0	Yes   No   Mo   No   No   No   No   No   No	mdents  to moderate with the m	1.0 0 1.0 0	phones	Provided   Provided	multiplelines Billing   phone service Electroni No Maile No Electron Yes Credit card ( phone service Yes Electron No Maile Yes Electron No Maile Yes Electron No Credit card No Electron	internet payment	DSL  er optic  DSL  er optic  pr optic	29.85  70.7  99.65  89.1  29.75  104.8  49.95  105.5  55.2  90.05	vity online   regs totalch   29.85    Yes    108.15    No    151.65    No    820.5    No    301.9    No    301.9    No    304.6.05    Yes    2686.05    No    5036.3    Yes    2686.05    No    1862.9	Backup di anges ch- yes  0  Yes  1  No  1  No  0  No  0  No  0  No  0  No  0  No  0  No  1  No  0  No  1  No  No  1  No  1  No  1  No  1  No  1  No  No  No  1  1  No  1  1  No  1
ustomerID gender si otection techSuppoi	Nol Nol Yes Nol Nol Nol Nol Nol Nol Nol		Tren'   depe     streamingM     Yes   No   Mo     No   No     No   Mo     No   Mo     No   Mo     No   Mo     No   Mo     No   Mo     Yes   M     Yes   M     No   Mo     Yes   M     No   Yes   M     Yes   M     No   No	Molinthion	1.00 tonth  2.00 tonth  4.00 tonth  4.00 tonth  2.00 t	phoneses	No No No Yess Yes Yes Yes Yes Yes Yes Yes Yes Ye	multiplelines Billing   phone service Electroni No Maile Relectroni Flectroni Selectroni Selectroni Maile Pass Flectroni No Maile Yes Flectroni No Maile Flectroni No Maile Selectroni No Maile Flectroni No No Flectroni Flectroni Flectroni Flectroni Flectroni Flectroni Flectroni	internet payment	DSL   Portice   Portice	onlineSec tonthlyCha 29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7  105.5  55.2  90.05	vesi Noi 3846.85 Noi 1862.9 Noi Noi Noi 1862.9 Noi 1862.9 Noi 1862.9 Noi Noi 1862.9 Noi Noi 1862.9 Noi	### ##################################
ustomerID gender so otection techSuppor 5590-VHYEG Female    No  668-0PYBK  Male    No  327-HQITU Female    No  327-HQITU Female    No  327-HQITU Female    No  3452-KIOVK  Male    No  822-POOKP Female    No  822-POOKP Female    No  280-XJGEX  Male    No  280-XJGEX  Male    No  280-XJGEX  Male    S  No  280-XJGEX  Male    S  Yes    190-MFLUM Female    S  Yes    S  Yes    190-MFLUM Female    S  Female    S  Female    S  Wale    No    173-30-MFLUM Female    S  Wale	Nol Nol Nol Yes  Nol Yes  Yes  Nol Yes	izen p mingTV		Molentification of the control of th	1.00   1.00	phoneses	No No No Yes  Yes  Yes  Yes  Yes  Yes  Yes  Yes	multiplelines Billing   phone service Electroni No Maile No Electroni Yes   Electron No Maile   Fleetron No Maile   Electron No Cledit card No   Credit card No	internet   payment   c check    d check    Fibe   c check    Fibe   ic check    Fibe   auto    Fibe   ic check    Fibe   Fibe   C check    C check    C check    C check    C check    C check	DSL  DSL  DSL  DSL  DSL  DSL  DSL  DSL	29.85  53.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7  105.5  55.2  90.05  39.65	vesi   S87.45  No  2986.3  Vesi   S87.45  No  528.35  No  1862.9  No  39.65	### ABackupi di narges   ch
ustomerID gender sc otection techSuppor S990-VHVEG Female	Nol Nol Yes Nol Nol Nol Nol Nol Nol Nol	izen pingTV 0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0	Tren'   depe     streamingM     Yes   No   Mo     No   No     No   Mo     No   Mo     No   Mo     No   Mo     No   Mo     No   Mo     Yes   Mo     No   Mo     Yes   Mo     No   Mo     Yes   Mo     No   Yes     No   Mo     Yes   Mo     Yes	Molentification of the control of th	1.00 ionth  2.00 ionth  2.00 ionth  2.00 ionth  2.00 ionth  2.00 ionth  3.00 ionth  4.00 ionth	phoneses	No No Vessi Yesi Yesi Yesi Yesi Yesi Yesi Yesi	multiplelines Billing   phone service Electroni No Maile Relectroni Flectroni Selectroni Selectroni Maile Pass Flectroni No Maile Yes Flectroni No Maile Flectroni No Maile Selectroni No Maile Flectroni No No Flectroni Flectroni Flectroni Flectroni Flectroni Flectroni Flectroni	internet payment c check    d check    Fibb ic check    Sibb ic check    Sibb	DSL   Portice   Portice	29.85  70.7  99.65  89.1  29.75  104.8  49.95  103.7  105.5  55.2  90.05  39.65	vesi Noi 3846.85 Noi 1862.9 Noi Noi Noi 1862.9 Noi 1862.9 Noi 1862.9 Noi Noi 1862.9 Noi Noi 1862.9 Noi	### Sackup[d args] ### Sackup[d args] ### Yes  ### 1   No  ###

以上就完成了模拟对电信公司的用户数据进行简单的预处理

接下来将采用 Kaplan-Meier 曲线来对数据进行生存分析。生存曲线以留存时间为横轴,其他特定数据指标为纵轴,绘制成连续型的阶梯形曲线,用以说明生存时间与生存率之间的关系。

从 silver 数据内获取 tenure 和 churn 并拟合 KM 模型。

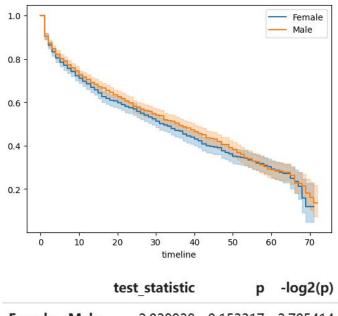
<Axes: title={'center': 'Kaplan-Meier Survival Curve: Population level'}, xlabel='timeline'>



#### 观察留存时间中位数:

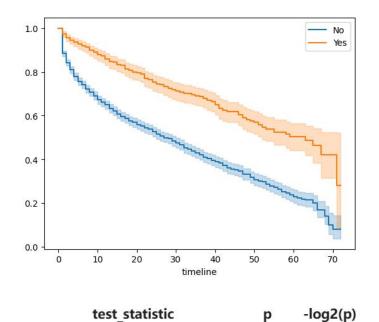
np.float64(34.0)

定义函数绘制在协变量水平上的 Kaplan-Meier curve 和打印 Log-rank 测试结果 有关性别和生存时间的关系:



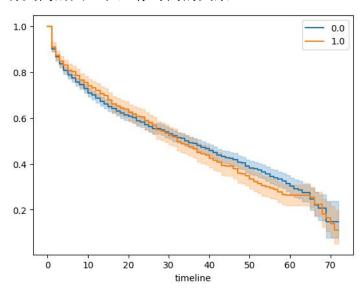
Female Male 2.038938 0.153317 2.705414

## 有关网络安全和生存时间的关系:



No Yes 141.60316 1.187554e-32 106.053706

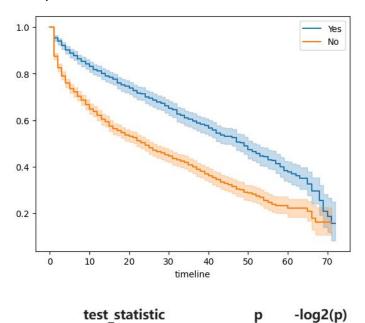
#### 有关高龄人士和生存时间的关系:



 test\_statistic
 p
 -log2(p)

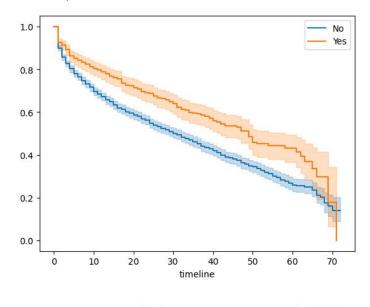
 0.0
 1.0
 0.125471
 0.723174
 0.467584

# 有关 partner 和生存时间的关系:



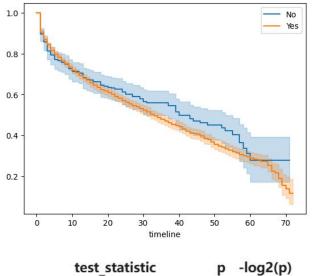
No Yes 135.758896 2.252911e-31 101.807981

# 有关 dependents 和生存时间的关系:



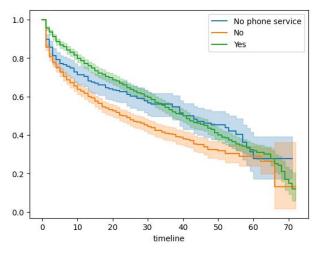
 No
 Yes
 35.031241
 3.244576e-09
 28.199323

# 有关 phoneService 和生存时间的关系:



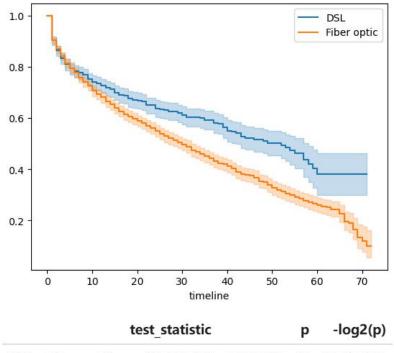
No Yes 1.683709 0.194432 2.36266

## 有关 multipleLines 和生存时间的关系:



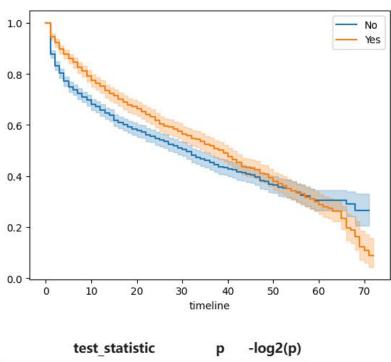
		test_statistic	р	-log2(p)
No	No phone service	12.382712	4.333273e-04	11.172255
	Yes	72.358368	1.794602e-17	55.629114
No phone service	Yes	1.500291	2.206266e-01	2.180322

## 有关 internetService 和生存时间的关系:



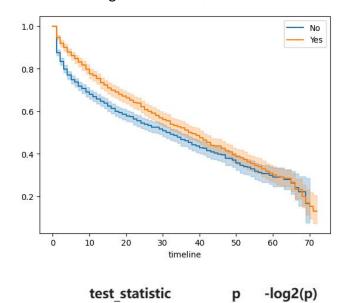
**DSL Fiber optic** 25.172866 5.241449e-07 20.863531

# 有关 streamingTV 和生存时间的关系:



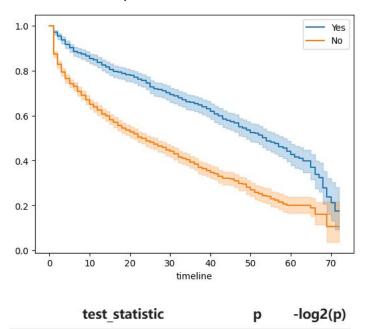
No Yes 12.93926 0.000322 11.601718

# 有关 streamingMovies 和生存时间的关系:



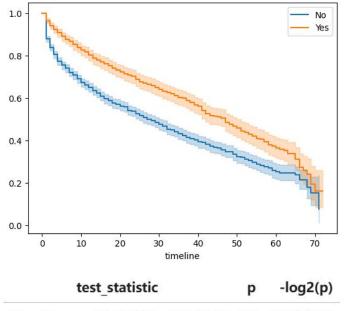
No Yes 17.941685 0.000023 15.422016

## 有关 onlineBackup 和生存时间的关系:



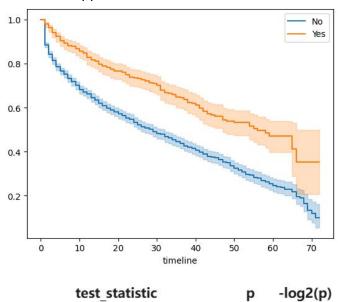
No Yes 189.482865 4.122979e-43 140.799221

## 有关 deviceProtection 和生存时间的关系:



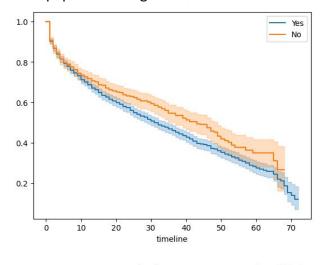
No Yes 71.496825 2.777047e-17 54.999226

# 有关 techSupport 和生存时间的关系:



No Yes 90.430334 1.916059e-21 68.822348

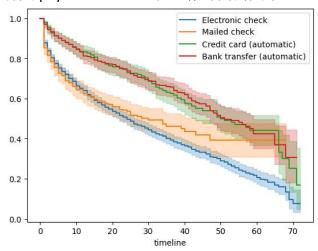
# 有关 paperlessBilling 和生存时间的关系:



 test\_statistic
 p
 -log2(p)

 No
 Yes
 8.340802
 0.003876
 8.011049

## 有关 paymentMethod 和生存时间的关系:



		test_statistic	р	-log2(p)
Bank transfer (automatic)	Credit card (automatic)	0.061543	8.040732e-01	0.314601
	Electronic check	91.191889	1.303937e-21	69.377616
	Mailed check	43.536998	4.160192e-11	34.484559
Credit card (automatic)	Electronic check	79.991082	3.761035e-19	61.205504
	Mailed check	39.684613	2.984678e-10	31.641706
Electronic check	Mailed check	0.898320	3.432326e-01	1.542741

在完成分析之后,可以提取生存概率以用于其他程序进行预测分析等。 定义一个函数用于获取特定生存率(以 DSL 为例):

#### DSL

- 0 1.000000
- **1** 0.902698
- 2 0.864380
- **3** 0.834702
- 4 0.810522
- **5** 0.794352
- 6 0.783900
- **7** 0.776362
- 8 0.768486
- 9 0.750833