

UTS Big Data Kelas B

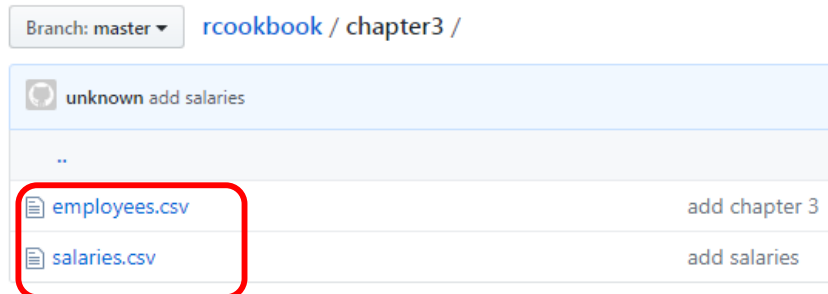
Pengerjaan tidak sesuai instruksi maka tidak akan dinilai

Instruksi:

1. Buatlah folder baru dengan nama: UTS_NIM pada halaman **desktop** komputer.
2. Buatlah **project baru** pada R studio dengan nama: UTS_KlsB_NIM, dan simpan pada folder baru (instruksi no 1)
3. Sifat ujian open akses, boleh tanya teman lewat internet (dilarang bertanya kepada teman secara langsung)

Soal:

1. Buatlah R script baru (ctrl+shift+N) kemudian simpan dengan nama: 1_NIM
 - a. Kunjungi halama ini: <https://github.com/ywchiu/rcookbook/tree/master/chapter3>
Download kedua file: employees.csv dan salaries.csv



Baca file tersebut dengan R, dan berilah nama: **karyawan** dan **gaji**. Hasilnya sebagai berikut:

V1	V2	V3	V4	V5	V6
10001	1953-09-02	Georgi	Facello	M	1986-06-26
10002	1964-06-02	Bezalel	Simmel	F	1985-11-21
10003	1959-12-03	Parto	Bamford	M	1986-08-28
10004	1954-05-01	Chirstian	Koblick	M	1986-12-01
10005	1955-01-21	Kyoichi	Maliniak	M	1989-09-12
10006	1953-04-20	Anneke	Preusig	F	1989-06-02
10007	1957-05-23	Tzvetan	Zielinski	F	1989-02-10
10008	1958-02-19	Saniya	Kalloufi	M	1994-09-15
10009	1952-04-19	Sumant	Peac	F	1985-02-18
10010	1963-06-01	Duanqkaew	Piveteau	F	1989-08-24

V1	V2	V3	V4
10001	60117	1986-06-26	1987-06-26
10001	62102	1987-06-26	1988-06-25
10001	66074	1988-06-25	1989-06-25
10001	66596	1989-06-25	1990-06-25
10001	66961	1990-06-25	1991-06-25
10001	71046	1991-06-25	1992-06-24
10001	74333	1992-06-24	1993-06-24
10001	75286	1993-06-24	1994-06-24
10001	75994	1994-06-24	1995-06-24
10001	76884	1995-06-24	1996-06-23
10001	80013	1996-06-23	1997-06-23

Simpan kedua file tersebut dalam bentuk excel dengan nama: **karyawan_NIM** untuk data karyawan, dan **gaji_NIM** untuk data gaji

- b. Download data excel dengan link:

<https://community.tableau.com/servlet/JiveServlet/downloadBody/1236-102-2-15278/Sample%20-%20Superstore.xls>

Baca file tersebut dengan R, dan berilah nama: **data_pesanan**. Hasilnya sebagai berikut:

Row.ID	Order.ID	Order.Date	Ship.Date	Ship.Mode	Customer.ID	Customer.Name	Segment	Country	
1	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United S
2	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer	United S
3	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United S
4	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United S
5	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United S
6	6	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United S
7	7	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United S
8	8	CA-2014-115812	2014-06-09	2014-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United S

Simpan file tersebut dalam bentuk excel dengan nama: **datapesanan_NIM**

2. Buatlah R script baru (ctrl+shift+N) kemudian simpan dengan nama: 2_NIM

- a. Berdasarkan data pada soal no 1 point b,

buatlah data baru dengan nama: **datapesanan_NIM**, dengan mengambil kolom 2,3,4,6,7,9,10,14,15,16,21.

- b. Rubahlah nama kolomnya dalam Bahasa Indonesia sehingga hasilnya sebagai berikut:

ID Pesanan	Tgl Pesanan	Tgl Pengiriman	ID Pemesan	Nama Pemesan	Negara	Kota	ID Produk	Kategori	Sub-Kategori	Keuntungan	
1	CA-2016-152156	2016-11-08	2016-11-11	CG-12520	Claire Gute	United States	Henderson	FUR-BO-10001798	Furniture	Bookcases	419136
2	CA-2016-152156	2016-11-08	2016-11-11	CG-12520	Claire Gute	United States	Henderson	FUR-CH-10000454	Furniture	Chairs	219.5820
3	CA-2016-138688	2016-06-12	2016-06-16	DV-13045	Darrin Van Huff	United States	Los Angeles	OFF-LA-10000240	Office Supplies	Labels	6.8714

- c. Buatlah kolom baru dengan nama **Lama Pesan** yang merupakan jarak antara **Tgl Pengiriman** dengan **Tgl Pesanan**

- d. Buatlah data baru yang memuat **ID Produk**, **Kategori**, **Sub-Kategori**, **Keuntungan** (0 atau minus), **Lama Pesan** (lebih dari 3 hari), sehingga data tersebut merupakan data penjualan produk yang tidak memiliki keuntungan dan proses kirimnya lebih dari 3 hari. Hasilnya sebagai berikut:

	ID Produk	Kategori	Sub-Kategori	Keuntungan	Lama Pesan
4	FUR-TA-10000577	Furniture	Tables	-383.0310	7
15	OFF-AP-10002311	Office Supplies	Appliances	-123.8580	4
16	OFF-BI-10000756	Office Supplies	Binders	-3.8160	4
28	FUR-BO-10004834	Furniture	Bookcases	-1665.0522	4
29	OFF-BI-10000474	Office Supplies	Binders	-7.0532	4
33	OFF-BI-10001525	Office Supplies	Binders	-5.7150	4
39	FUR-BO-10002545	Furniture	Bookcases	-46.9764	4
40	FUR-CH-10004218	Furniture	Chairs	-15.1470	4
67	FUR-CH-10001146	Furniture	Chairs	-15.2225	5
73	FUR-CH-10000513	Furniture	Chairs	-114.3912	6
75	OFF-ST-10004123	Office Supplies	Storage	-18.1960	6
79	FUR-FU-10003194	Furniture	Furnishings	-14.4750	5
96	OFF-BI-10004738	Office Supplies	Binders	-3.7880	6
101	TEC-AC-10001767	Technology	Accessories	-10.7973	4
102	OFF-BI-10002609	Office Supplies	Binders	-3.0396	4
104	TEC-AC-10001552	Technology	Accessories	-26.8758	4
105	FUR-FU-10004006	Furniture	Furnishings	-3.8385	4
106	OFF-BI-10002794	Office Supplies	Binders	-25.8174	4

Simpan data baru tersebut dalam bentuk excel dengan nama: **produkruji_NIM**

3. Buatlah R script baru (ctrl+shift+N) kemudian simpan dengan nama: 3_NIM
 - a. Download data “purchase_order” dan “purchase_view” pada link: <https://github.com/ywchiu/rcookbook/tree/master/chapter6>
 - b. Buatlah data yang memuat jumlah barang yang dilihat dan yang dibeli secara keseluruhan. Buatlah kolom baru yang memuat prosentase barang dibeli dengan hitungan: (barang dibeli/barang dilihat)x100. [kolom frekuensi barang dibeli tidak memunculkan angka 0 atau NA]

Product	Frek_dibeli	Frek_dilihat	prosentase_dibeli
P0000005913	3	14	21.428571
P0000006020	1	13	7.692308
P0000006591	1	3	33.333333
P0000007744	2	13	15.384615
P0000008190011	1	0	Inf
P0000008352	1	4	25.000000
P0000010194	1	2	50.000000
P0000010706	1	5	20.000000
P0000011163	2	21	9.523810
P0000011270	1	3	33.333333
P0000011362	1	9	11.111111
P0000011421	1	9	11.111111
P0000011524	1	1	100.000000
P0000013506	1	7	14.285714

Simpan data tersebut dalam bentuk excel dengan nama: **prosentaseproduk_NIM**

- c. Buatlah data yang memuat jumlah produk yang dipesan user tiap tanggal, sehingga dapat diketahui perilaku user pada tanggal-tanggal tertentu membeli produk apa saja

Tgl 1 - 30

User	Product	1	2	3	4	5	6	7	8	9	10	11	12
U1000001354	P0014520085												
U10000044910	P0018800250	-	-	-	-	-	-	-	-	-	-	-	-
U10000056953	P0003926850034	1	-	-	-	-	-	-	-	-	-	-	-
U10000056953	P0013344760004	-	-	-	-	-	-	-	-	-	-	-	-
U10000056953	P0013834251	-	-	-	-	-	-	-	-	-	-	-	-
U10000056953	P0014251480003	-	-	-	-	-	-	-	1	-	-	-	-
U10000065461	P0023597022	-	-	-	-	-	-	-	-	-	-	-	-
U10000092625	P0018474750044	-	-	-	-	-	-	-	-	-	-	-	-
U10000092625	P0023729451	-	-	-	-	-	-	-	-	-	-	-	-
U10000092625	P0024077600013	-	-	-	-	-	-	-	-	-	-	-	-
U10000092625	P0024236730	-	-	-	-	-	-	-	-	-	-	-	-
U10000110348	P0003440476	-	-	-	-	-	-	-	-	-	-	1	-
U10000118351	P0003169854	-	-	-	-	-	-	-	-	-	-	-	-

Simpan data tersebut dalam bentuk excel dengan nama: **tanggalbeli_NIM**

- d. Buatlah data yang memuat perilaku user terhadap product tiap tanggal, sehingga bisa diketahui pada tanggal tertentu seorang user melihat berapa kali suatu produk, tapi tidak membeli produk tersebut. [kolom frekuensi barang dilihat min. 10, dan kolom frekuensi barang dibeli nilainya 0]

User	Time	Product	Frek_dilihat	Frek_dibeli
U1000092625	2015-08-08	P0004598882	12	0
U10001550955	2015-07-29	P0023492453	10	0
U10002121866	2015-08-12	P0025776461	19	0
U10003515768	2015-07-31	P0006993652	10	0
U10003535535	2015-08-14	P0000145364	10	0
U10004605734	2015-08-26	P0005661913	13	0
U10004605734	2015-08-26	P0022273635	10	0
U10004605734	2015-08-26	P0024251474	13	0
U10006639140	2015-08-08	P0021911470	10	0
U1000789075	2015-08-30	P0009962260	11	0
U1000789075	2015-08-30	P0009969083	34	0
U10008182069	2015-07-31	P0018765154	10	0
U1000905389	2015-08-28	P0005449651	11	0
U10009629366	2015-08-13	P0025279730	11	0
U10010400073	2015-08-08	P0025223376	16	0
U10012347769	2015-08-08	P0000075036	10	0
U10013300448	2015-07-31	P0022618400	10	0
U10016880847	2015-08-29	P0022852233	12	0
U1001694551	2015-07-31	P0005484286	11	0

Simpan data tersebut dalam bentuk excel dengan nama: **perilakuuser_NIM**