Group Assignment Open Source Programming 2022

Prof. Dr. Matthijs Meire

Introduction:

You will receive data about an online gambling website. The company is interested in creating knowledge from their databases; however, they are not really sure how to do this. Your task is to provide 1) basic marketing descriptives for the customer base and 2) create a marketing data mart per customer (so a row for each customer and columns contain key marketing metrics for that customer).

Data:

See the documentation file on Github for a detailed description of the data. Be sure that you are familiar with the data before starting the assignment.

Goals:

Your goal is to create marketing insights from these data. You are going to create a 'marketing data mart', in which you capture the important marketing metrics for the betting company. These marketing data are on a customer level, so you will have to create a basetable with one row per customer. The columns should be the different marketing metrics. Examples of marketing metrics are loyalty, consumption, …

Next to this, we want an overall picture on the customers. E.g., where do the customers come from, what is their average age, what type of products do customers buy in specific regions, etc. We want this in a flexible and interactive format, using an R Shiny application. Make sure to be creative and use nice visuals where possible!

(You do not have to do predictive or prescriptive analyses, but make detailed descriptive analyses)

Output:

- Your R scripts, with comments so that this is understandable
- Create a neat and concise Marketing data manual for marketing analysts, that contains the overall summary statistics that you think are relevant and also gives a detailed description of the gamblers' metrics calculated. Do this in an easy, reproducible way.
- Your data mart and overall summary statistics
- A Shiny application, which provides interactive features to investigate the marketing numbers

Due dates:

- Mid-term: November 5

- o Provide the marketing data mart, which is a basetable summarizing the data per customer (see above)
  - o Provide the code you used to read in and process your data
  - o 25% of the grade
- Final: 2 weeks after the final session (December 20 at latest)

Codebook for Actual Internet Sports Gambling Activity:

February 2005 through September 2005

This codebook provides information for the raw datasets used to generate the first epidemiological report of actual Internet sports betting (LaBrie et al., 2007). These datasets come from the collaborative Internet gambling research project between the Division on Addictions (DOA) and bwin Interactive Entertainment, AG (bwin), an Internet betting service provider headquartered in Vienna, Austria. These datasets provide evidence from the first eight months of the first prospective longitudinal, real-time, Internet sports betting behavior study that took place from February 1, 2005 through September 30, 2005.

Source(s): Division on Addictions, Cambridge Health Alliance, a teaching affiliate of Harvard Medical School

PI(s): Dr. Howard J. Shaffer

Sponsor(s): bwin, Interactive Entertainment, AG

Related publication(s): LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of Internet sports gambling behavior. Journal of Gambling Studies, 23, 347-362.

Note: the appendixes are added as an Excel file as well for easier use in R.

Raw Dataset I - Demographics

The raw dataset, Demographics, contains the following demographic information: User ID, Country of Residence, Language, Gender, Registration Date, First Pay-in Date, First Active Date, First Sports Book Active Date, First Casino Active Date, First Games Active Date, First Poker Active Date, and Application ID. Information about City of Residence and Date of Birth are censored to protect participant privacy and comply with institutional review board (IRB) requirements.

Raw Dataset II - UserDailyAggregation

The raw dataset, UserDailyAggregation, contains the actual betting information associated with each product (please see appendix 1) for each participant for each calendar day with at least one transaction from February 1, 2005 through September 30, 2005. This dataset contains the following information: User ID, Date, Betting Product, Stakes, Winnings, and Bets. All transactions have been converted to Euros.

Raw Dataset III - PokerChipConversions

The raw dataset, PokerChipConversions, contains the actual poker chip transaction information from February 1, 2005 through September 30, 2005. The DOA obtained summaries of poker play in terms of poker chip transactions to and from the poker site for each poker session. The poker transaction information includes User ID, Transaction Date/Time, Transaction Type (buy or sell), and Transaction Amount.

Data mart

For the data mart, aggregate betting behavior data for bwin Internet sports subscribers who opened an account with bwin during the period from February 1, 2005 through February 27, 2005. Participants who had no missing values for first active date and who registered during the period of February 1, 2005 through February 27, 2005 should be entered into the data mart (i.e., in the raw dataset Demographics, variable FirstAct does not equal missing and variable RegDate ranges from February 1, 2005 through February 27, 2005).

You can exclude records in the raw dataset UserDailyAggregation that took place before the first pay-in date (i.e., variable FirstPay in raw dataset Demographics) in the preparation of the data mart. . The exclusion controlled for activities financed by promotional money that might not reflect the participants' behavior when personal funds are at stake. First active date of each product (not including poker) can be identified from the raw dataset UserDailyAggregation, and first poker active date can be identified by first date of poker chip purchase from raw dataset PokerChipConversions. Therefore, first active date can be assigned as the earliest date of the first active date of all betting products.

Variables in Dataset 1 Demographics

Variable	Date Type	Description	N	Min	Max	Note
UserID	Numeric	UserID	42649	1324354	1402198	User ID was assigned to each participant by bwin at time of registration.
Country	Numeric	Country of residence	42649	8	895	Please see appendix 2.
Language	Numeric	Primary Language	42649	1	17	Please see appendix 3.
Gender	Numeric	Gender	42649	0	1	0: female; 1:male
RegDate	Text	Registration Date	42649	Feb-01-2005	Feb-27-2005	Date of registration
FirstPay	Text	First Pay-in date	42649	Feb-01-2005	Oct-02-2005	Participant's first betting money deposits date.
FirstAct	Text	First Active Date	42647	Feb-01-2005	Oct-02-2005	Participant's first active play date.
FirstSp	Text	First Sports Book play date	41413	Feb-01-2005	Oct-02-2005	Participant's first sports book play date
FirstCa	Text	First Casino play date	5888	Feb-01-2005	Oct-02-2005	Participant's first casino play date
FirstGa	Text	First Games play date	3273	Feb-01-2005	Oct-02-2005	Participant's first games play date
FirstPo	Text	First Poker play date	2142	Feb-01-2005	Oct-02-2005	Participant's first poker play date
ApplicationID	Numeric	Betting application	42649	1	42	Different route of access to bwin. Please see appendix 4.

Variables in Dataset 2 UserDailyAggregation

Variable	Data Type	Description	N	Min	Max	Note
UserId	Numeric	UserID	1740196	1324354	1405190	User ID was assigned to each participant by bwin at time of registration.
Date	Text	Date of Betting Activities Aggregation	1740196	Feb-01-2005	Sep-30-2005	Date of betting activity aggregation.
ProductId	Numeric	Betting product ID	1740196	1	8	Different betting product each participant betted on. Please see appendix 1.
Stakes	Numeric	Total Betting Money (euro)	1740196	-28.3	272415.5	Total betting stakes on a given day (Euro). The negative numbers was due to the accounting system correction.
Winnings	Numeric	Total Winnings Credited to Participant (Euro)	1740196	-40	270222.0	Total betting winnings on a given day (Euro). The negative numbers was due to the account system correction.
Bets	Numeric	Total Number of bets	1740196	-1	6223	Total number of bets on a given day. The negative numbers was due to the accounting system correction.

Variables in Dataset 3 PokerChipConversions

Variable	Data Type	Description	N	Min	Max	Note
userId	Numeric	User ID	275480	1324355	1402198	User ID was assigned to each participant by bwin at time of registration.
TransDateTime	Text	Poker Chip Transaction Date and Time	275480	Feb-01-2005	Sep-30-2005	Poker Chip Transaction Date and Time
TransType	Numeric	Poker Chip Transaction Type	275480	24	124	124 : sell ; 24 : buy
TransAmount	Numeric	Poker Chip Transaction Amount (Euro)	275480	0.000064437	15995.53	Poker Chip Transaction Amount (Euro)

Appendix 1. Codings for productID

ProductID	Product Description
1	Sports book fixed-odd
2	Sports book live-action
3	Poker BossMedia
4	Casino BossMedia
5	Supertoto
6	Games VS
7	Games bwin
8	Casino Chartwell

Appendix 2. Codings for country

Country	Country Name
4	Afghanistan
8	Albania
10	Antarctica
12	Algeria
16	American Samoa
20	Andorra
24	Angola
28	Antigua and Barbuda
31	Azerbaijan
32	Argentina
36	Australia
40	Austria
44	Bahamas

48	Bahrain
50	Bangladesh
51	Armenia
52	Barbados
56	Belgium
60	Bermuda
64	Bhutan
68	Bolivia
70	Bosnia and Herzegovina
72	Botswana
74	Bouvet Island
76	Brazil
84	Belize
86	British Indian Ocean Territory
90	Solomon Islands
92	Virgin Islands (British)
96	Brunei Darussalam
100	Bulgaria
104	Myanmar
108	Burundi
112	Belarus
116	Cambodia
120	Cameroon
124	Canada
132	Cape Verde
136	Cayman Islands
140	Central African Republic
144	Sri Lanka
148	Chad
152	Chile

156	China
158	Taiwan
162	Christmas Island
166	Cocos (Keeling) Islands
170	Colombia
174	Comoros
175	Mayotte
178	Congo
180	the Democratic Republic of the Congo
184	Cook Islands
188	Costa Rica
191	Croatia
192	Cuba
196	Cyprus
203	Czech Republic
204	Benin
208	Denmark
212	Dominica
214	Dominican Republic
222	El Salvador
226	Equatorial Guinea
231	Ethiopia
232	Eritrea
233	Estonia
234	Faroe Islands
238	Falkland Islands
242	Fiji
246	Finland
250	France

254	French Guinea
258	French Polynesia
260	French Southern Territories
262	Djibouti
266	Gabon
268	Georgia
270	Gambia
276	Germany
288	Ghana
292	Gibraltar
296	Kiribati
300	Greece
304	Greenland
308	Grenada
312	Guadeloupe
316	Guam
320	Guatemala
324	Guinea
328	Guyana
334	Heard and Mc Donald Islands
336	Vatican City State (Holy See)
340	Honduras
344	Hong Kong
348	Hungary
352	Iceland
356	India
360	Indonesia
364	Iran
368	Iraq
372	Ireland

376	Israel
380	Italy
384	Cote d'Ivoire
388	Jamaica
392	Japan
398	Kazakhstan
400	Jordan
404	Kenya
408	Korea, Republic of
410	Korea (South)
414	Kuwait
417	Kyrgyzstan
418	Lao Peoples Democratic Republic
422	Lebanon
426	Lesotho
428	Latvia
430	Liberia
434	Libya
438	Liechtenstein
440	Lithuania
442	Luxembourg
446	Macau
450	Madagascar
454	Malawi
458	Malaysia
462	Maldives
466	Mali
470	Malta
474	Martinique
478	Mauritania

480	Mauritius
484	Mexico
492	Monaco
496	Mongolia
498	Moldavia
500	Montserrat
504	Morocco
508	Mozambique
512	Oman
516	Namibia
520	Nauru
524	Nepal
528	Holland
530	Netherlands Antilles
533	Aruba
540	New Caledonia
548	Vanuatu
554	New Zealand
558	Nicaragua
562	Niger
566	Nigeria
570	Niue
574	Norfolk Island
578	Norway
580	Northern Mariana Islands
581	Unites States Minor Outlying Islands
583	Micronesia, Federated States of
584	Marshall Islands
585	Palau

586	Pakistan
591	Panama
598	Papua New Guinea
600	Paraguay
604	Peru
608	Philippines
612	Pitcairn
616	Poland
620	Portugal
624	Guinea-Bissau
626	East Timor
630	Puerto Rico
634	Qatar
638	Reunion
642	Romania
643	Russian Federation
646	Rwanda
654	St. Helena
659	Saint Kitts and Nevis
660	Anguilla
662	St. Lucia
666	St. Pierre and Miquelon
670	Saint Vincent and the Grenadines
674	San Marino
678	Sao Tome and Principe
682	Saudi Arabia
686	Senegal
690	Seychelles
694	Sierra Leone
702	Singapore

703	Slovakia
704	Vietnam
705	Slovenia
706	Somalia
710	South Africa
716	Zimbabwe
724	Spain
732	Western Sahara
736	Sudan
740	Suriname
744	Svalbard and Jan Mayen Islands
748	Swaziland
752	Sweden
756	Switzerland
760	Syria
762	Tajikistan
764	Thailand
768	Togo
772	Tokelau
776	Tonga
780	Trinidad and Tobago
784	United Arab Emirates
788	Tunesia
792	Turkey
795	Turkmenistan
796	Turks and Caicos Islands
798	Tuvalu
800	Uganda
804	Ukraine
807	FYR Macedonia

818	Egypt
826	United Kingdom
834	Tanzania
840	USA
850	United States Virgin Islands
854	Burkina Faso
858	Uruguay
860	Uzbekistan
862	Venezuela
876	Wallis and Futuna Islands
882	Samoa (Independent)
887	Yemen
891	Serbia and Montenegro
894	Zambia
895	Diego Garcia
896	Ascension
897	Australian overseas territories
898	Tokelau
1000	Undefined

Appendix 3. Codings for language

Language	Language Description
1	English
2	German
3	Italian
4	Spanish
5	Swedish
6	French

7	Turkish
8	Greek
9	Polish
10	Norwegian
11	Danish
12	Catalan
13	Czech
14	Hungarian
15	Dutch
16	Portuguese
17	Russian
18	Slovenian
19	Croatian
20	Slovak
21	Simple Chinese
22	Traditional Chinese

Appendix 4. Codings for applicationID

ApplicationID	Application Description
1	BETANDWIN.COM
2	TRIPLE-A-CASINO.COM
3	BETANDWIN.DE
4	WAP.BETANDWIN.COM
5	SMS MOBILE BETTING APPLICATION
6	CHARITY
7	DOLCEVITACASINO.COM
8	BALLS OF FIRE

9	BETEUROPE.COM	
10	BAHSEGIR.COM	
11	CASINO.BETEUROPE.COM	
12	WWW.CASINOTURK.COM	
13	WWW.SANALCASINO.COM	
14	BETOTO.COM	
15	PLAYIT.COM	
16	CASINO.PLAYIT.COM	
17	THECROUPIER.COM	
18	SMS.BETANDWIN.COM	
19	WAP.BETANDWIN.DE	
21	BOF.PLAYIT.COM	
22	BOF.BETEUROPE.COM	
23	BETANDWIN POKER	
24	BETANDWIN CASINO	
26	SMS.BETANDWIN.DE	
27	LOTTERY.BETOTO.COM	
28	BETWORK.COM	
29	WAP.PLAYIT.COM	
30	PLAYIT POKER	
31	BETEUROPE POKER	
32	BETOTO POKER	
33	BETANDWIN GAMES	
36	BETOTO CASINO	
38	BETEUROPE GAMES	
42	PLAYIT GAMES	