5 **Solo 1 Assignment: The App Happy Company and
the Market for Social Entertainment Apps**

The App Happy Company wants to better understand what they believe is the market for a new social entertainment app they are thinking of developing. They are currently in the
10 business of providing B2B analytic apps, and don't have yet have a product in the consumer entertainment app category.

App Happy hired the Consumer Spy Corporation (CSC) to survey consumers in the market. CSC collected data from a sample of consumers, and provided App Happy with a dataset of
15 of their responses. The survey questionnaire (it's in the file **apphappy-quex-sp2016.pdf**) was based on some preliminary qualitative research that included focus groups and one-on-one interviews.

(1) App Happy wants you to use their survey data to do a *general attitudinal post hoc*
20 *segmentation analysis*. You should develop and evaluate a segmentation scheme for App Happy, profile the segments in your scheme, interpret your results, and make recommendations about product opportunities or additional research as might be appropriate. Be sure to describe your analytic methodology and to summarize assumptions, caveats and limitations that App Happy should bear in mind in using what you report to them. Based on
25 the requirements provided to CSC, App Happy assumes that the data are reasonably representative of the market of interest.

Note that the purpose of this segmentation analysis is to highlight opportunities to develop and execute segment-specific marketing strategies and tactics. Your results should include
30 targeting recommendations. *Be sure to try and to report on using at least two different clustering algorithms.* Explain why you chose to use the algorithms that you applied.

See below for additional "Guidance" on doing this assignment.

35 (2) App Happy would also like to use your results to put consumers that it doesn't currently have data on into the segments you will define for them should it be able to obtain data on them. It would like you to describe how you would develop a classification model that could do this. You don't have to actually apply the model to get results. *You just need to describe what model or models they could use, and provide enough explanation that App Happy would*
40 *be well informed to apply the model(s) you recommend by itself*.   In the application domain of segmentation analysis, such models are sometimes called "typing tools."   (But they have nothing to do with keyboards...)

You may use additional information or data about the market as seems useful or important to
45 you.

Be sure to check the data for errors or anomalies as a preliminary step in your segmentation

analysis. Also, consider whether data transformations may be useful. Note that the data are real data and the documentation on them is less than perfect. Finally, **be sure to address all**
50   **the above issues and questions**

App Happy's data are in the R data file **apphappyData-sp2016.Rdata**. This file contains to R data frames: **apphappy.4.num.frame,** and **apphappy.4.labs.frame**. The first of these has numerically coded survey response data.   The second has response data coded in the
55   character strings of the questionnaire's value labels. These strings are like the value labels in SPSS, or the labels you might find attached to numerical codes in SAS when Proc Format is used.

Guidance
60

<u>For (1):</u>

This is a *post hoc descriptive market segmentation analysis* exercise. Analyses of this type have two general parts: (a) Use data to identify groups, that differ on what are called "basis"
65   variables, and (b) "profile" the groups using data to highlight unique characteristics so as to inform segment-specific strategies and tactics.

To do (a) you need to decide what data to use.   App Happy wants an analysis based on attitudes, so you need to use data from questionnaire items that are about attitudes. (Hint:
70   the Likert scaled items are attitudinal measures.) The attitudinal item variables will be your basis variables, the variables with which you will try to find groups that might comprise useful attitudinal segments. You'll use them as well as the other data available from App Happy to define and gain understanding of your segments' unique characteristics as they might apply to marketing strategy.
75
To find your groups you'll use one or more *cluster analysis algorithms.* There are many such algorithms. Common approaches include hierarchical clustering, and partitioning methods that include k-means clustering.   More advanced methods include model-based clustering, ensemble clustering, and clustering based on copulas. We'll talk about some of these in our
80   Sync sessions. The R cluster task view available on your preferred R mirror summarizes the many different clustering methods and tools available to you. (We're using R in this course, right?)

It's important to note that all methods require making some assumptions about the nature of
85   your basis variables. An important consideration is whether your basis variables are discrete or continuous measures.   Discrete measures are either nominal (unordered categories) or ordinal (ordered) categories.   It doesn't make mathematical sense to apply operations like summing or averaging discrete variables. It is sometimes the case that discrete data or mixtures of discrete and continous data are transformed into a matrix of distances that is then
90   used as data for clustering. Gower's Coefficient is one of the ways this transformation to distances (or similarities, kind of the complement of distances) is accomplished.

Given that different algorithms apply different methods to the data, it's almost always useful to use more than one clustering method when doing segmentation analysis.   A main goal is to

95   define well-separated clusters that differ in ways that are relevant to marketing efforts. Since in post hoc analysis you don't know a priori how many segments there might be, you need to examine and compare clustering solutions in order to pick out a solution. Well separated clusters have little overlap. Techniques for assessing cluster separation include comparing within vs. between variances, using information theory-based measures like the Bayesian

100   Information Criterion (BIC) for model based clustering solutions, and employing graphical methods like the Silhouette Plot. If you review what's in the R Cluster Task View and do a little online searching you'll see that many different approaches are available for your use. In R, see for example the R package clValid.

105   Note that visualizing clusters in high dimensions is in general very difficult unless they they have simple, convex shapes, and they span just a few (say two or three) dimensions.

How many clusters can you expect to find? In general, anywhere from zero to a very large number. The basic idea is to let the data tell you how many there are. With survey data like

110   App Happy's, it'd be unlikely to find more than a handful, well separated and cohesive clusters. Strategic segmentation studies based on survey data will generally end up producing no more that seven at most. With true behavioral or online data, you might find many more. Such segmentations are usually used for more tactical purposes than for strategy development. Online retailers and ad servers may rely on hundred's of segments.

115

Don't forget: your goal for (1)(a) is to come up with groups that you feel are distinct in terms of their attitudes. App Happy wants you to do a *general attitudinal post hoc segmentation analysis*

120   (1)(b): Profiling is about describing how your segments differ, and in particular about how is one is uniquely different from the others. A segment's "profile" is a summary of the important ways it differs from other segments with respect to marketing.

You can use your basis variables and the other variables in App Happy's data to profile your

125   segments. Good practice consists of making *statistical* comparisons to do this so that the differences you highlight are likely to be reliable. Just comparing points estimates (e.g. sample means) is not ideal because between them may just reflect random error. Be aware of the "experimentwise error rate" is you are testing many hypotheses about differences between your segments on your basis or auxilliary (non-basis) variables. You will want to

130   consider whether you should be using a technique for adjusting for the effects of making multiple comparisons.

Graphical methods can be usefu when profiling segments. The "parallel plot" is one such technique. They can be done in various R packages. Try searching online for "R CRAN

135   parallel plot" to find options.

Part (2):

A typing tool is a predictive classifier. This assignment asks that you recommend one for App

140   Happy to use, and that you explain your recommendation. There are many different kinds of classifiers. They include among others multinomial logit or probit regression models,

classification tree models, CHAID models, support vector machines (SVMs), and RandomForest. The common element is that they predict an unordered, discrete outcome variable, which in the present case would be membership in the segments you will have

145   defined for AppHappy.   Your task for (2) is to look into these methods and to select one that AppHappy might use. You'll find your effort worthwhile as you'll be using predictive classifers for Solo 3.

Be sure to take a look at supplementary materials provided on Canvas, and to review the

150   required and recommended readings.