Homework 1 - due 23:59, 27/Sep/2018

Problem 1 (40 points). Use bike dataset (bike.xlsx) provided in Moodle to segment customers based on bikes purchased by applying the k-means clustering algorithm. Analyze customer segments and provide short conclusion based on results. Those are the steps you need to take:

1. Transform the dataset to an applicable data for analysis representing customers in each row, products (bikes) in column and quantities bought as values,
2. Run k-means algorithm and create optimal number of clusters of customers, save the modified data to an excel file.
3. Use Excel to create a final dataset and analyze it (using VLOOKUP/INDEX-MATCH, pivot tables, conditional formatting and other formulas and features if necessary).
4. Submit final Python file/notebook and the final excel sheet. Short conclusion of analysis made should be submitted in a separate word/pdf document together with answers to Problem 3 and 4.

Problem 2 (30 points). Coefficients of the regression model ran on ice-cream survey data is provided in the ice.xlsx. The attributes are ice-cream Flavor (5 levels), Packaging (3), Light (2) and Organic (2). Use the already analyzed data (i.e. coefficients) to calculate the relative importance of each attribute and understand what is the highest utility bundle (i.e. what are the levels per attribute that provide highest utilities). Your answers should be provided in the 2nd sheet of the excel file.

Problem 3 (20 points). Tashir Pizza decided to estimate customer loyalty using NPS. The following table provides results for Branch A (Northern Avenue, Yerevan) and B (Vanadzor):

	Detractors (%)	Neutrals (%)	Promoters (%)
Branch A	20	40	40
Branch B	33	33	34

Given the information above, Tashir Pizza management concludes that customers in Branch A (Northern Avenue, Yerevan) are very loyal.

Clearly state whether you agree or not with the conclusion and provide 3 grounded reasons to justify your position.

Problem 4 (10 points). Choose one of the options for each question.

Correlation refers to:

1. the causal relationship between two variables,
2. the association between two variables,
3. the proportion of variance that two variables share,
4. all of the above.

Conjoint analysis may be employed to:

1. identify important attributes that influence consumer choice,
2. estimate the probability of customer churn,
3. identify customer segments,
4. all of the above.