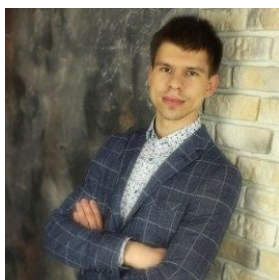


УДК 004.62+338.48

СТАТИСТИЧЕСКИЙ АНАЛИЗ СООБЩЕСТВ СОЦИАЛЬНЫХ СЕТЕЙ: ПОТРЕБИТЕЛЬСКИЕ ИССЛЕДОВАНИЯ



И.А. Труханович
Магистрант кафедры ПОИТ
БГУИР,
инженер-программист



А.И. Парамонов
Доцент кафедры ПОИТ БГУИР,
кандидат технических наук, доцент

Белорусский государственный университет информатики и радиоэлектроники,
Республика Беларусь
E-mail: ilya.trukhanovich@gmail.com

И.А. Труханович

Окончил Белорусский государственный университет информатики и радиоэлектроники в 2018 году. Магистрант БГУИР. Работает в EPAM Systems в должности инженера-программиста.

А.И. Парамонов

Доцент кафедры программного обеспечения информационных технологий БГУИР.

Аннотация. Для принятия любого решения в рамках предложения услуг важно понимать ситуацию на рынке, учитывать возможности, выявить тенденции и механизмы. Таким образом, для достаточной эффективности в данном вопросе необходимо проведение исследований рынка. Одним из инструментов подобного рода исследований может стать сбор актуальной статистики тематических сообществ в социальных сетях, а затем её последующий анализ. В работе предлагается подход получения и обработки статистики из социальных сетей с языка статистической обработки R. Показаны возможности используемых пакетов для применения в потребительском исследовании.

Ключевые слова: социальные сети, потребительское исследование, пакеты языка R, туризм.

Введение. Туризм является одним из крупнейших бизнес-секторов мировой экономики. Экономическое значение отрасли хорошо иллюстрируется тем фактом, что за последние 10 лет на туризм приходилось ежегодно порядка 10% мирового ВВП. Вклад туризма в глобальную экономику в 2017 году превысил 8 триллионов долларов США — 10,4% от общего объема мирового ВВП, а уже в 2018 объем туристического рынка достиг рекордного значения в \$8,8 трлн. Роль туризма и его влияние на мировую экономику подтверждается данными обзора Всемирного совета по туризму и путешествиям (WTTC). Исследование WTTC [1] выявило, что туристическая отрасль обеспечивает каждое десятое рабочее место в мире: в 2018 году таких мест было 319 млн, а к 2028 году количество вырастет до 414 млн. По прогнозам организации, к 2028 году вклад туризма в мировой ВВП составит уже \$12,5 трлн или 11,7% от общего объема. В Республике Беларусь туризм является одним из приоритетных направлений социально-экономического развития. Данная отрасль взята под особую опеку государства. Сегодня в Беларуси реализуется ряд госпрограмм и приняты необходимые указы [2].

Одним из самых распространенных видов предоставления туристических услуг являются туристские предприятия или организации, которые занимается разработкой маршрутов и туров (турфирмы). В турфирмы постоянно обращается множество граждан с просьбой подобрать им тур. Порой такие запросы являются настолько специфичными, что не представляется возможным наличие сразу сформированных вариантов ответов. Однако в большинстве случаев клиенты не очень расположены к длительному ожиданию расчета их маршрута и предпочитают выбрать из уже имеющихся пакетов. С целью поддержания интереса и спроса турфирмам необходимо ориентироваться на потребности клиентов. Этап определения потенциальных потребностей клиента подразумевает анализ рынка, опросы и исследования. Для эффективного потребительского исследования необходимо обладать достаточной совокупностью релевантных и актуальных данных. Поставщиком информации подобного рода может служить статистика соответствующих тематических сообществ в социальных сетях, которая, как правило, располагается в открытом доступе. Интернет, и, в частности, социальные сети, сегодня предоставляют большие объемы информации, однако при работе с ними возникают серьезные задачи по извлечению релевантных достоверных данных и последующему их анализу. Одним из важных этапов потребительского исследования становится выбор инструмента анализа данных. В данной работе для получения статистических данных и последующего их анализа предлагается использовать язык статистической обработки **R** [3].

Пример получения данных. Для демонстрации возможностей среды **R** по сбору и анализу статистических данных в сети Интернет выбрано одно из сообществ в социальной сети «ВКонтакте», которое связано с обсуждением туристических направлений в русскоязычном пространстве. Данные в группе отобраны за период с 2016 по 2019. Доступ к API «ВКонтакте» реализован с помощью пакета *rvkstat* [4]. Визуализация геолокации была выполнена с помощью сервиса Google Maps. Доступ к API Google Maps реализован с помощью пакета *ggmap* [5].

Пример скрипта для получения статистики посетителей сообщества, в том числе с учётом городов и возрастов, отображен на рисунке 1.

```
communityData <- vkGetGroupStat(date_from = "2016-01-01", date_to = "2019-12-09", group_id = groupId, access_token = vkToken$access_token)
communityCitiesData <- vkGetGroupStatCity(date_from = "2016-01-01", date_to = "2019-12-09", group_id = groupId, access_token = vkToken$access_token)
communityAgeData <- vkGetGroupStatAge(date_from = "2016-01-01", date_to = "2019-12-09", group_id = groupId, access_token = vkToken$access_token)
```

Рисунок 1. – Запросы на получение статистики посетителей сообщества

Отображение географии посетителей. Для отображения географии посетителей была выполнена группировка по городам, после чего с помощью данных Google Maps Api получены координаты для каждого из городов. С помощью сервиса Google Maps Api была загружена карта Европы. Данные посетителей сообщества перед визуализацией предварительно были отфильтрованы – из данных по посетителям были убраны неевропейские города, а сами данные разделены на три группы согласно количеству. На полученную карту были нанесены данные по посетителям, а результат сохранён в файл. Пример скрипта, который выполняет визуализация статистики посетителей на карте Европы, представлен на рисунке 2.

```
png('EuropeMap.png')
EuropeMap +
  geom_point(aes(x=lon, y=lat), data=aggregatedCommunityEuropeCitiesDataMax,
  col="red", alpha=1, size=aggregatedCommunityEuropeCitiesDataMax$x*0.0001) +
  scale_size_continuous(range=range(aggregatedCommunityEuropeCitiesDataMax$x
  )) +
  geom_point(aes(x=lon, y=lat),
  data=aggregatedCommunityEuropeCitiesDataAverage, col="orange", alpha=1,
  size=aggregatedCommunityEuropeCitiesDataAverage$x*0.0001) +
  scale_size_continuous(range=range(aggregatedCommunityEuropeCitiesDataAvera
  ge$x)) +
  geom_point(aes(x=lon, y=lat), data=aggregatedCommunityEuropeCitiesDataMin,
  col="black", alpha=1, size=aggregatedCommunityEuropeCitiesDataMin$x*0.0001) +
  scale_size_continuous(range=range(aggregatedCommunityEuropeCitiesDataMin$x))
  dev.off()
```

Рисунок 2. – Нанесение данных на карту

Пример результата работы скрипта отображён на рисунке 3.

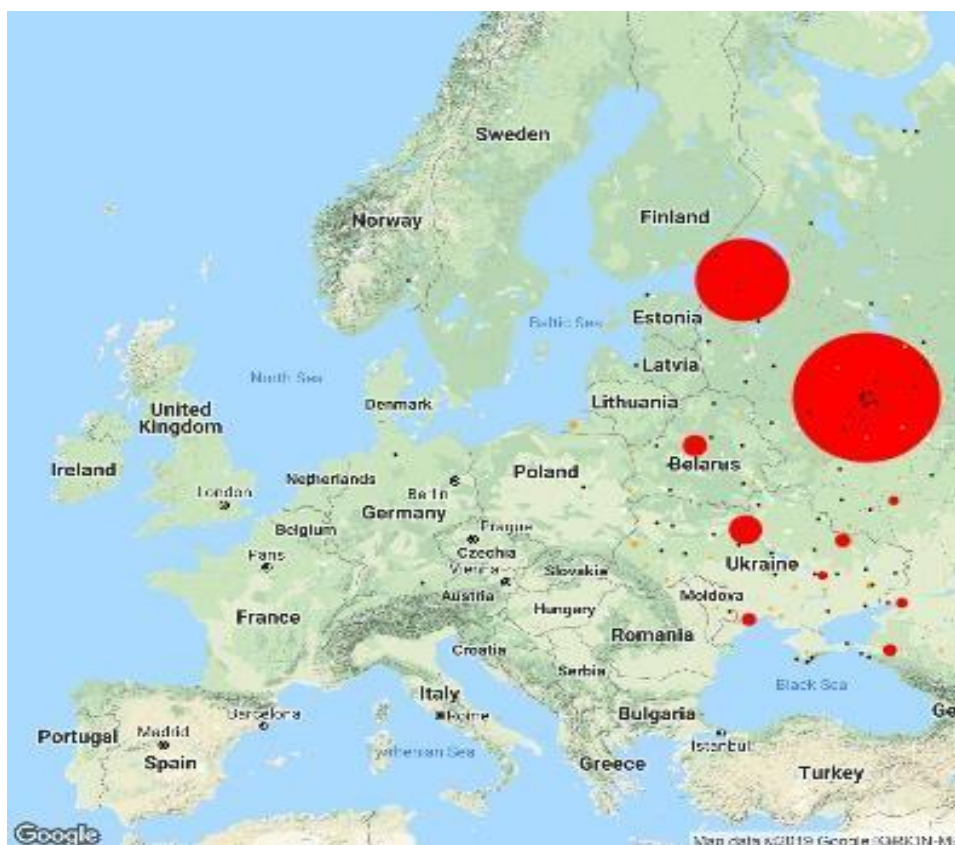


Рисунок 3. – География посетителей сообщества

Отображение статистики подписчиков. Язык R обладает мощными средствами для построения различных графиков и визуализации данных. Одним из таких инструментов является пакет *lattice* [6]. С его помощью возможна визуализация статистики, позволяющая сделать вывод о количестве посетителей с течением времени в возрастном разрезе. Результат выполнения функции *xypplot* для отображения статистики посетителей по возрасту представлен на рисунке 4.

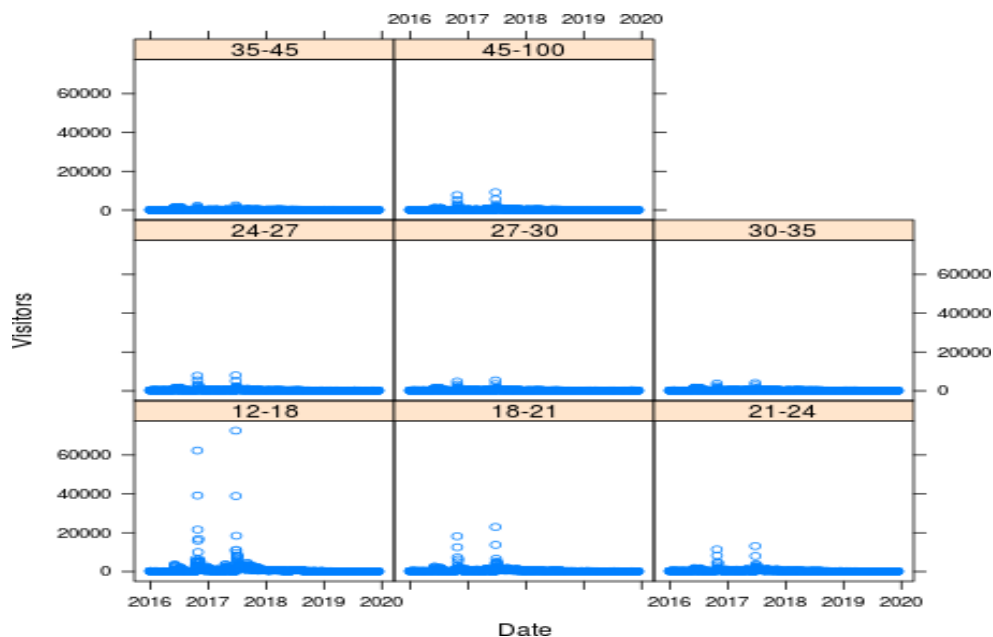


Рисунок 4. – Статистика посетителей по возрасту

Пакет *forecast* [7] содержит инструменты для визуализации временных рядов. С его помощью можно отобразить сезонную динамику роста количества подписчиков. Результат применения функции *ggseasonplot* к данным сообщества представлен на рисунке 5. Как видно из графика, с мая 2017 года наблюдается динамика роста, снижающая свои показатели в периоды последних месяцев последующих лет.

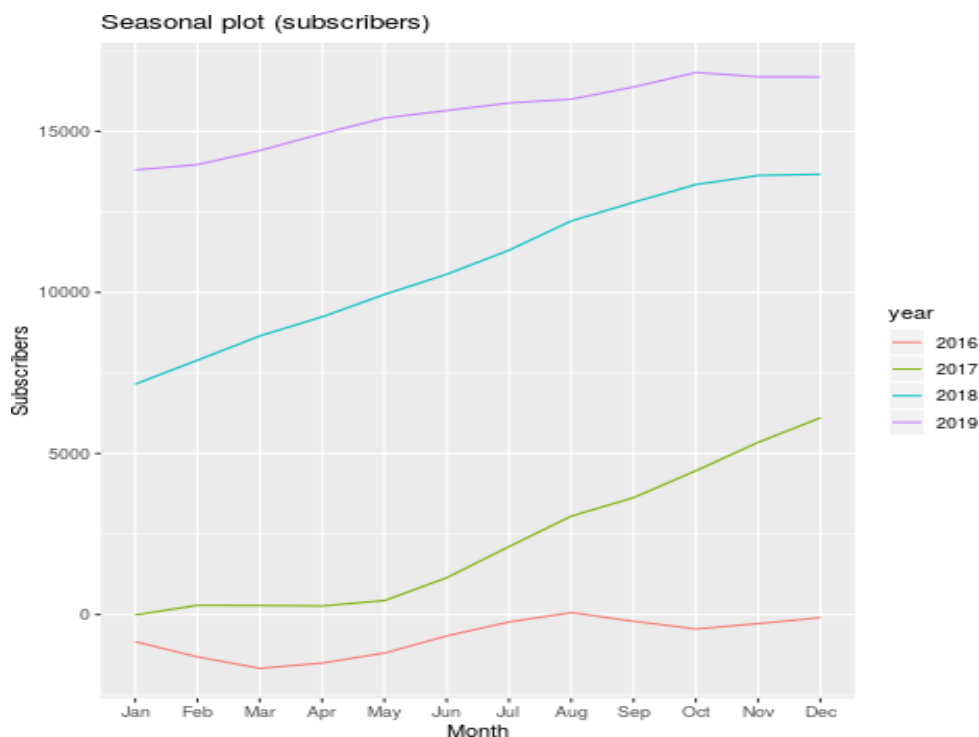


Рисунок 5. – Сезонный график роста подписчиков

Интерпретация полученных результатов. Как сказано ранее, вышеприведённый подход можно использовать в рамках исследования целевой аудитории туристической компании. В частности, можно применять данные методы анализа в вопросах изучения спроса туристических маршрутов. Предложенный подход можно применять для исследования нескольких тематических сообществ, посвящённых туристическим городам, объектам и т.д. После этого полученную статистику можно обобщить, выполнить анализ в совокупности и определить сезонную динамику спроса, возрастной диапазон, динамику спроса и прочее.

Список литературы

- [1.] World Travel and Tourism Council [Electronic resource]. – Mode of access: <https://www.wttc.org>. – Date of access: 20.01.2019.
- [2.] Официальный интернет-портал Президента Республики Беларусь [Электронный ресурс]. – Режим доступа: http://president.gov.by/ru/tourism_ru. – Дата доступа: 25.01.2019.
- [3.] Matloff, N. The Art of R programming: A Tour of Statistical Software Design / N. Matloff. – San Francisco : No Starch Press, 2011. – 441 p.
- [4.] Package rvkstat [Electronic resource]. – Mode of access: <https://www.rdocumentation.org/packages/rvkstat>. – Date of access: 25.01.2019.
- [5.] Package ggmap [Electronic resource]. – Mode of access: <https://www.rdocumentation.org/packages/ggmap>. – Date of access: 26.01.2019.
- [6.] Package lattice [Electronic resource]. – Mode of access: <https://www.rdocumentation.org/packages/lattice>. – Date of access: 26.01.2019.
- [7.] Package forecast [Electronic resource]. – Mode of access: <https://www.rdocumentation.org/packages/forecast>. – Date of access: 26.01.2019.

STATISTICAL ANALYSIS OF SOCIAL NETWORKS COMMUNITIES IN CONSUMER RESEARCH ISSUES

I.A. TRUKHANOVICH

*Master student of Software for
Information Technologies
Department of BSUIR,
software engineer*

A.I. PARAMONOV,

*PhD (Candidate of Computer Sciences),
Associate Professor at Software for
Information Technologies Department of
BSUIR, Associate Professor*

Belarusian State University of Informatics and Radioelectronics, Republic of Belarus

E-mail: ilya.trukhanovich@gmail.com

Abstract. It is important to understand the market situation, take into account opportunities, identify trends and mechanisms to make any decision as part of the service offer. Thus, for sufficient efficiency in this matter, market research is necessary. One of the tools of this kind of research can be the collection of relevant statistics of thematic communities in social networks, and then its subsequent analysis. The paper proposes an approach to obtaining and processing statistics from social networks from the language of statistical processing R. The possibilities of the packages used for use in consumer research are shown.

Keywords: social networks, consumer research, R language packages, tourism.