



Научный семинар по исследованиям цифровой экономики «Цифровизация и демография»

3 февраля 2021 года

(по материалам НИР ЭФ МГУ «ВОСПРОИЗВОДСТВО НАСЕЛЕНИЯ В ЦИФРОВОМ ОБЩЕСТВЕ»)

Калабихина И.Е. (ЭФ МГУ, рук.)

Архангельский В.Н. (ЭФ МГУ)

Николаева У.Г. (ЭФ МГУ)

Колотуша А.В. (ЭФ МГУ, аспирант кафедры народонаселения)

Абдуселимова И.А. (ЭФ МГУ, магистр ЭП)

Банин Е.П. (МГТУ им. Н.Э. Баумана, НИЦ «Курчатовский институт», аспирант)

Клименко Г.А. (ЭФ МГУ, магистр ЭП)

Шамсутдинова В.Ш. (ЭФ МГУ, аспирантка кафедры народонаселения)

<https://demography.econ.msu.ru/about/science/infographics/> – проект 2 (инфографика на сайте кафедры)



Научный семинар по исследованиям цифровой экономики
«Цифровизация и демография»

3 февраля 2021 года

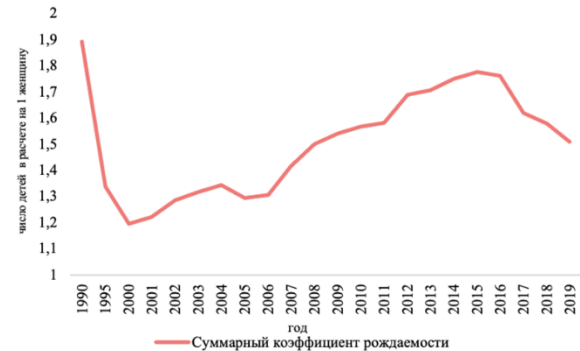
(по материалам НИР ЭФ МГУ «ВОСПРОИЗВОДСТВО НАСЕЛЕНИЯ В ЦИФРОВОМ ОБЩЕСТВЕ»)

О влиянии цифровизации на демографическое поведение и
демографические исследования

Калабихина И.Е.

Мотивация

- Значительный скачок роста доступа к высокоскоростному интернету в РФ (с 2013 по 2019 от 61% до 81% населения РФ, ЕМИСС)
- Отсутствие консенсуса/достаточных знаний по поводу влияния интернета на самосохранительное и репродуктивное поведение людей + наличие вызовов
- Рост объема больших данных
- Большой потенциал для демографической науки и политики



2 исследования в одном

Цель 1 «Оценка влияния доступа к интернету на демографическое поведение»

- Выявить влияние доступа к широкополосному интернету на репродуктивное и самосохранительное поведение на основе эконометрического моделирования по данным российских выборочных обследований

Цель 2 «Алгоритм использования больших данных в анализе демографического поведения»

- Разработать направления использования больших данных в демографическом анализе и политике (область интереса – рождаемость, репродуктивное поведение, мнения о демографической политике и пр.)

Что удалось получить к настоящему моменту

ЧАСТЬ 1

- *определена степень влияния доступа к широкополосному интернету на репродуктивное поведение,*
- *определена степень влияния использования широкополосного интернета на самосохранительное поведение,*

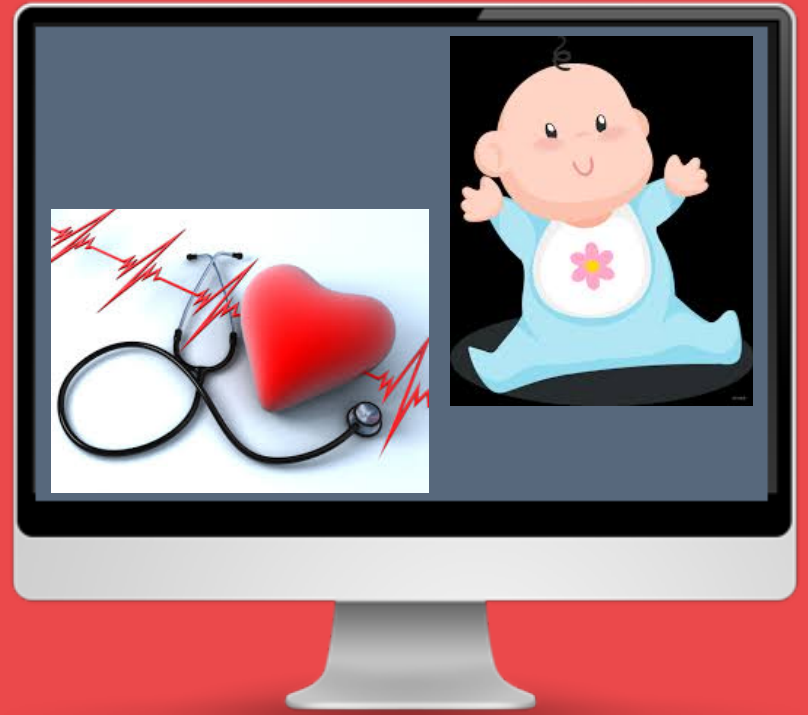
ЧАСТЬ 2

- *выполнены замеры интенсивности реакции пользователей разных групп на разные демографические политические меры в различных социальных сетях и поисковых системах - ВКонтакте, Google trend, Youtube, Yandex,*
- *разработана и апробирована методология измерения демографической температуры (тональности текстов, эмоционального фона) комментариев выбранных социально-демографических групп в социальной сети ВКонтакте,*
- *разработана и апробирована методология краткосрочного прогнозирования рождаемости по интенсивности запросов в поисковой системе Google trend*

Статьи и база данных

1. *Калабихина И. Е., Абдуселимова И. А., Клименко Г. А.* Влияние высокоскоростного интернета на репродуктивное поведение в России // *Вестник Московского университета. Серия 6: Экономика.* 2020. № 6. С. 90-103.
2. *Калабихина И. Е., Колотуша А. В.* Является ли Интернет сберегающим здоровье фактором в России? // *Демографическое обозрение.* 2020. Т. 7, № 3. С. 150-182. [DOI](#)
ПОБЕДА НА МОЛОДЕЖНОЙ СЕССИИ РЭК-2020 (Абдуселимова, Клименко, Колотуша)
3. *Kalabikhina IE, Banin EP.* Database “Pro-family (pronatalist) communities in the social network VKontakte” // *Population and Economics.* 2020. Vol. 4, no. 4. P.98-103
[10.3897/popecon.4.e60915](https://doi.org/10.3897/popecon.4.e60915) – ОТКРЫТАЯ БАЗА ДАННЫХ
4. *Калабихина И.Е., Банин Е.П., Абдуселимова И.А., Архангельский В.Н., Клименко Г.А., Колотуша А.В., Николаева У.Г., Шамсутдинова В.Ш.* Краткосрочное прогнозирование демографических тенденций на основе данных Google trends // *Прикладная информатика.* 2020. Т. 15, № 6. С. 91-118. [DOI](#)

ЧАСТЬ 1



О влиянии интернета на демографическое поведение

репродуктивное

- влияет на рост рождаемости в развитой стране (Германия) [Billari et. al., 2019]
- помогает совмещать профессиональную и семейную жизнь [Dettling, 2017]
- увеличивает ЗП высококвалифицированного работника и, следовательно, альтернативные издержки времени [Akerman et al., 2015]
- информирует о методах контрацепции и последствиях родительства за счет дополнительных запросов в интернете, что снижает рождаемость среди подростков в развив-ся странах [Guldi, Herbst, 2017]
- является новым эффективным способом поиска партнера [Rosenfeld, Thomas, 2012], позволяет снизить издержки поиска партнера и увеличить скорость поступления предложений знакомства [Bellou, 2015]; увеличивает общение с партнером [Wajcman et al., 2008], снижает общение [Boswell, Olson-Buchanan, 2007; Nie, 2001]

самосохранительное

- интернет и здоровье: просвещение о вреде курения и потребления алкоголя [Killian, 2012];
- распространение услуг для ведения здорового образа жизни [Feil et al., 2003; Swartz et al., 2006];
- длительное использование гаджетов без перерывов (усталость, искривление позвоночника, заболевание глаз и пр.) [Калабугина, 2013];
- пропаганда нездорового образа жизни [Груздева, 2017]

Результат 1. Влияние доступа к интернету на репродуктивное поведение позитивное

- Увеличивает вероятность рождения ребенка у женщин средних и старших возрастов (25-49 лет)
- Положительно влияет на решение о рождении второго и последующих детей, хотя мы находим слабый эффект и на рождение первого ребенка
- Положительно влияет на совмещение семейных и профессиональных обязанностей (баланс семья-работа)

Калабихина И. Е., Абдуселимова И. А., Клименко Г. А. Влияние высокоскоростного интернета на репродуктивное поведение в России // Вестник Московского университета. Серия 6: Экономика. 2020. № 6. С. 90-103.

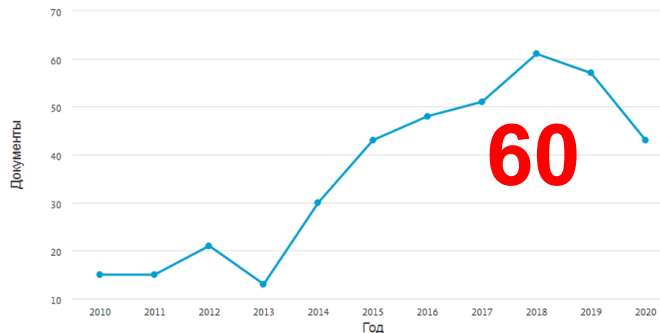
Результат 2. Влияние частоты использования интернета на самосохранительное поведение и здоровье преимущественно позитивное

- Женщины и мужчины, чаще использующие интернет, в среднем лучше оценивают своё здоровье
- Мужчины, которые чаще пользуются интернетом, при прочих равных в среднем меньше курят, меньше употребляют алкоголь в целом, замещают употребление водки употреблением вина
- Женщины, которые чаще пользуются интернетом, при прочих равных в среднем больше курят и употребляют алкоголь
- Дополнительно: образование, число детей, доход, состояние в браке положительно влияют на самооценку здоровья; проживание в городе и возраст – отрицательно; число детей оказывает сдерживающее влияние на потребление табака (особенно у мужчин)

Калабихина И. Е., Колотуша А. В. Является ли Интернет сберегающим здоровье фактором в России? // Демографическое обозрение. 2020. Т. 7, № 3. С. 150-182. [DOI](#)

ЧАСТЬ 2

Динамика публикаций по теме
«big data» в демографии



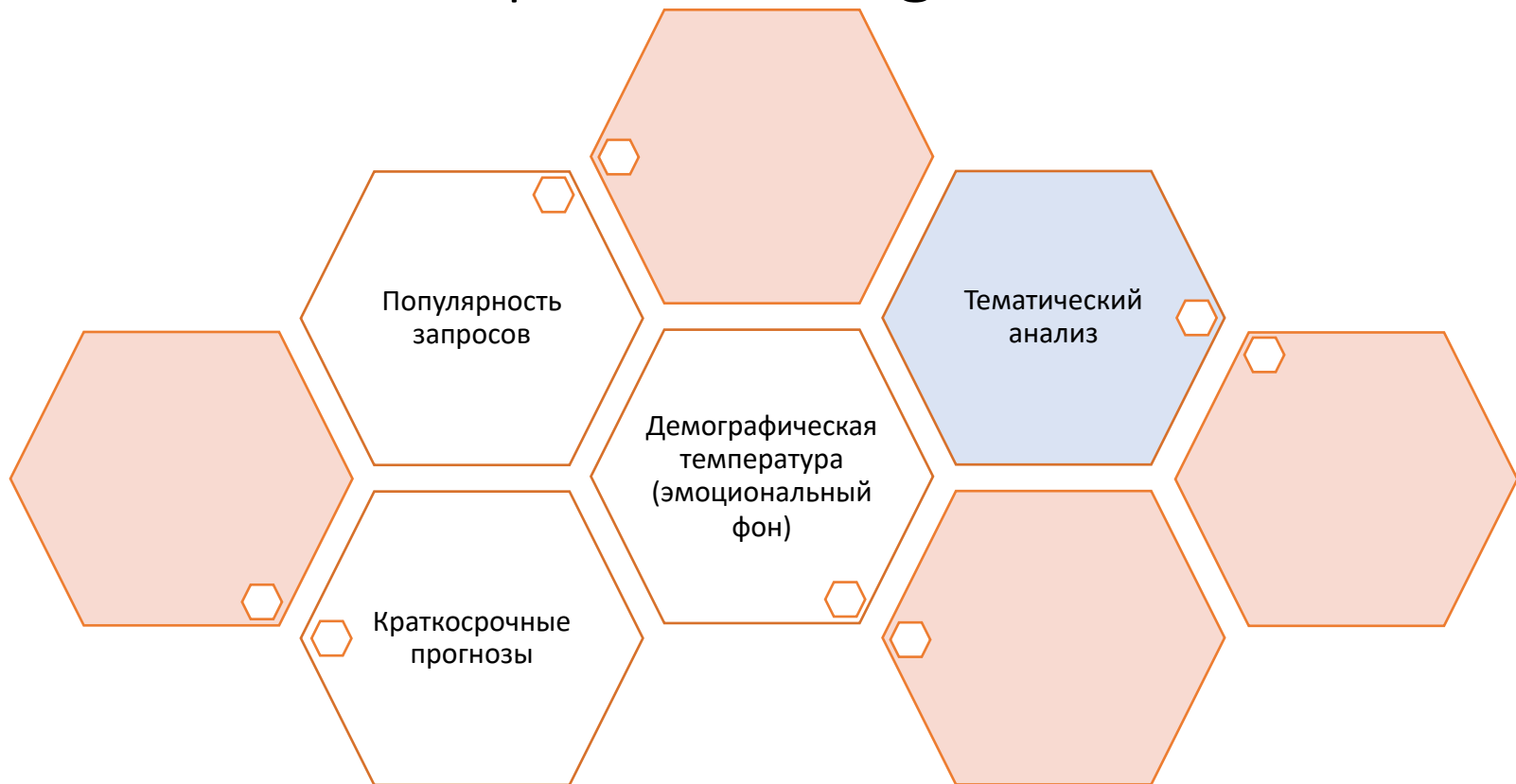
Июнь 2020



Социальные сети в социально-демографических исследованиях

- Twitter для изучения процессов расовой сегрегации [*Cesare et al 2017*].
- Facebook для изучения особенности ассимиляции мексиканских иммигрантов в США [*Stewart et al 2019*].
- Данные послужных списков (career histories) из резюме участников профессиональной социальной сети LinkedIn использовались в работе [*State et al 2014*] для определения объема миграционного притока образованных кадров и студентов в США и страны Азии.
- Google – нелегальные аборты в США [*Reis, Brownstein 2010*].
- Прогнозные модели рождаемости на короткий и средний горизонт прогнозирования (до двух лет) [*Billari 2016*].
- Большая часть работ посвящена анализу тональности высказываний с помощью лингвистического анализа текста в сети Twitter [*Miller, Dickinson, Hu 2012; Mitchell et al 2013*], Youtube [*Filippova 2015*]. В частности, анализ на основе информации из Twitter в Италии при использовании вычислительных лингвистических методов [*Mencarini et al 2019*] позволяет узнать, какие чувства испытывают люди будучи родителями, как планируют рождение детей.

Направления нашей работы с Big DATA



Результат 3. Популярность запросов и политика, и поведение

Непосредственная реакция пользователей на меры демографической политики выявляется относительно редко. Оценка влияния демографической политики на репродуктивное поведение (отношение, мнение) различных социально-демографических групп в России на основе данных из социальных сетей возможна по резонансным мерам (например, материнский капитал).

Изменения в популярности запросов, связанных с самим демографическим поведением дает дополнительные результаты

Популярность запросов

период: 2004 - 2020

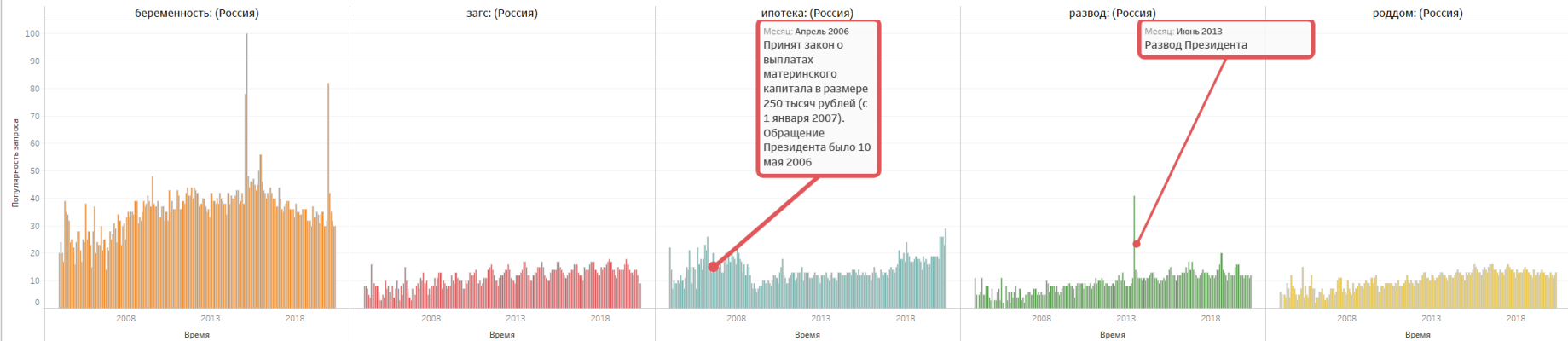
запрос: аборт, беременность, контрацепция, материнский капитал



Популярность запросов

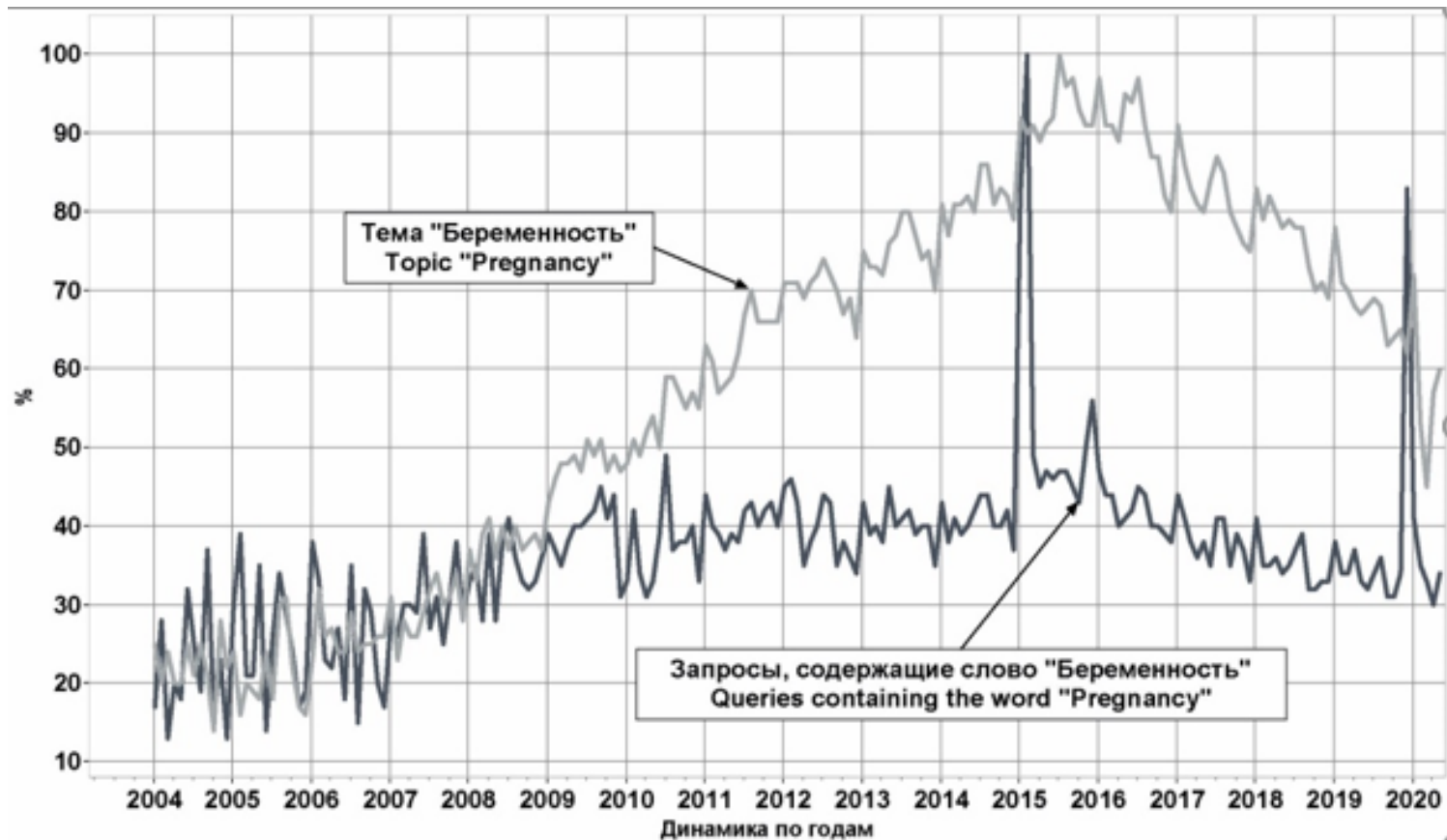
период: 2004 - 2020

запрос: беременность, загс, ипотека, развод, роддом



Темы использовать лучше, чем ключевые слова

Анализ текстовых запросов к Google Trends, 2009 - 2020



Результат 4. Анализ тональности текстов позволяет измерить демографическую температуру (эмоциональный фон) в определенных сегментах (социально-демографических группах) социальных сетей

Создана база данных «*Просемейные (пронаталистские) и **антинаталистские** сообщества в социальной сети ВКонтакте*»

Отрицательная демографическая температура (абсолютные значения) - социальные сети фиксируют преимущественно негативные эмоции. Измерение глубины отрицательной температуры - инструмент измерения факторов эмоционального фона в области репродуктивного поведения, политики, мнений

Kalabikhina IE, Banin EP. Database “Pro-family (pronatalist) communities in the social network VKontakte” // Population and Economics. 2020. Vol. 4, no. 4. P.98-103
[10.3897/popecon.4.e60915](https://doi.org/10.3897/popecon.4.e60915)

I.E.Kalabikhina, I.A.Abduselimova, E.P. Banin, G.A. Klimenko, A.V. Kolotusha. The measurement of demographic temperature using the tonality analysis of data from the social network Vkontakte // Mathematics. 2021 Forthcoming

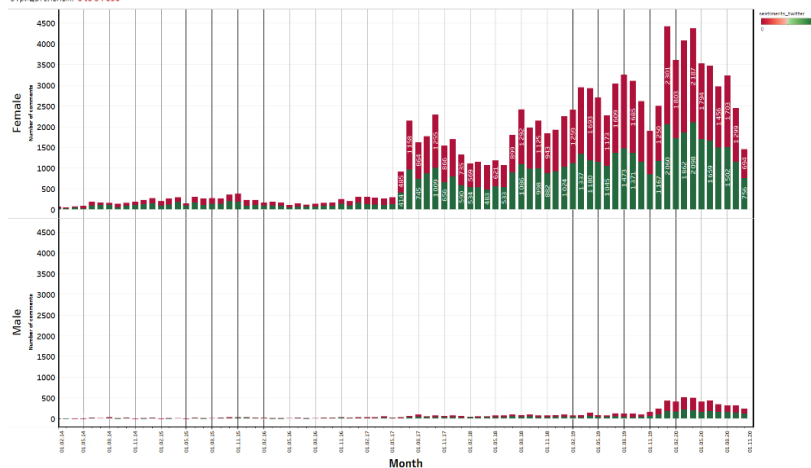
Распределение количества комментариев в корпусах данных по месяцам 02.2014-11.2020

«Красный» - негативные комментарии, «зеленый» - позитивные
Верхняя часть – женщины, нижняя - мужчины

Пронаталистские группы,
314 групп, ≈100 тыс. комментариев

Анализ тональности комментариев по месяцам (пронаталистские группы)

Положительных: 0 to 48 885
Отрицательных: 0 to 54 690

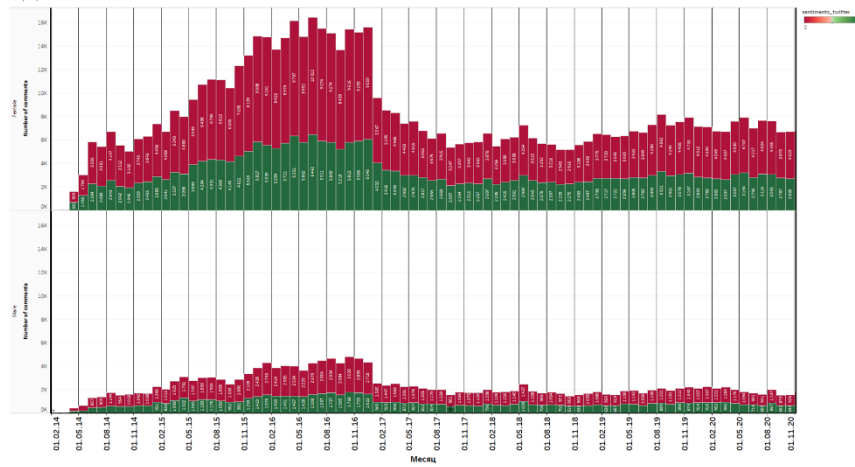


The list of source data files for data from 01.02.2014 to 01.11.2020 is located in the file: data_source_data_pro_natalist_groups.csv. The chart shows details about sentiment by gender. The chart is based on sentiment by gender, which excludes null.

Антинаталистские группы,
9 групп, ≈ 670 тыс. комментариев

Анализ тональности комментариев по месяцам (антинаталистские группы, полная выборка)

Положительных: 0 to 267 002
Отрицательных: 0 to 403 938

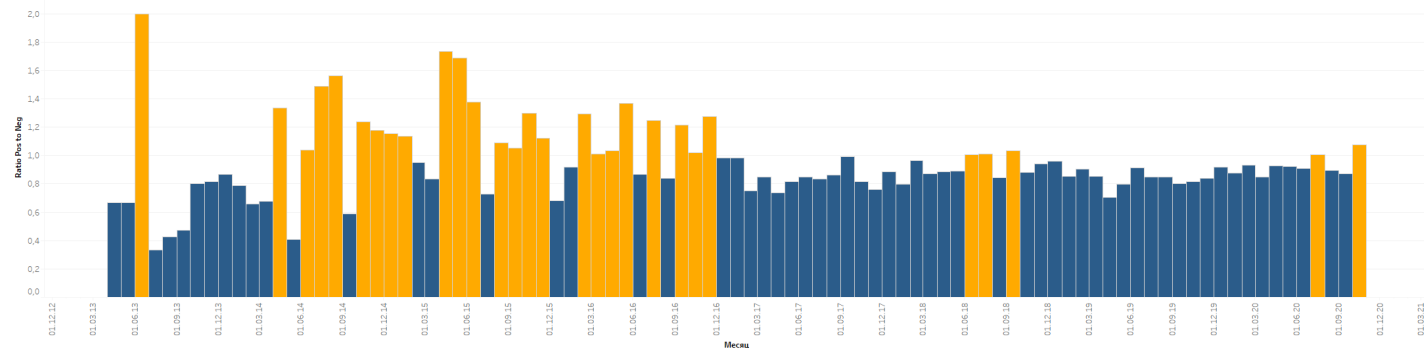


The list of source data files for data from 01.02.2014 to 01.11.2020 is located in the file: data_source_data_pro_natalist_groups.csv. The chart shows details about sentiment by gender. The chart is based on sentiment by gender, which excludes null.

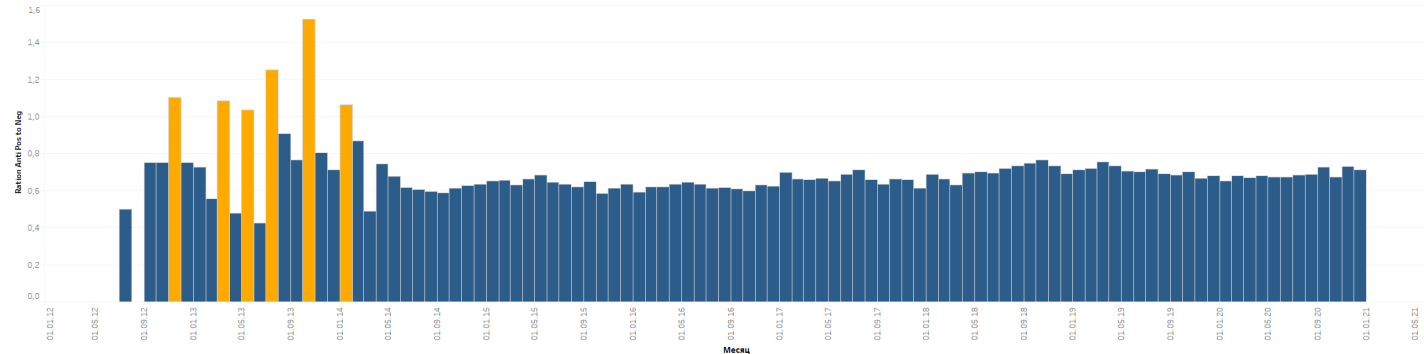
Демографическая температура (относительная)

Желтая +, синяя -. Вверху – пронаталисты, внизу – антинаталлисты. 2012-2020

Отношение положительных комментариев к отрицательным (пронаталисты)
по месяцам



Отношение положительных комментариев к отрицательным (антинаталлисты)
по месяцам



Результат 5. Существуют темы в анализе текстов поисковых запросов о семейных ценностях, рождении детей, родительстве, интенсивность запросов которых будет устойчиво коррелировать с данными официальной демографической статистики в области рождаемости, брачности и разводимости

Мы смогли на данном этапе разработать алгоритм краткосрочного прогнозирования рождаемости

Калабихина И.Е., Банин Е.П., Абдуселимова И.А., Архангельский В.Н., Клименко Г.А., Колотуша А.В., Николаева У.Г., Шамсутдинова В.Ш. Краткосрочное прогнозирование демографических тенденций на основе данных Google trends // *Прикладная информатика*. 2020. Т. 15, № 6. С. 91-118. [DOI](#)

Краткосрочное прогнозирование демографических тенденций на основе данных Google Trends

1. Разработан алгоритм поиска прокси-переменных к данным статистики, которые могут быть использованы в моделях множественной регрессии, для заполнения «пропусков» в демографических данных, для прогнозов
2. Разработана методика прогнозирования тенденций в области рождаемости в краткосрочном периоде на основе запросов к Google trend (помесячная рождаемость, годовая рождаемость, с лагами)
3. Апробирована методика выполнения краткосрочных прогнозов рождаемости в России (2009-2022). Ошибка прогноза на уровне страны 1,14-2,11% (на уровне региона высокая)



Перспективы применения больших данных в демографическом анализе существенны

- тематический анализ «демографических» мнений
- сравнение демографической температуры различных социальных и демографических групп населения
- углубленное изучение тональности на основе детализации эмоциональной окраски текстов
- изучение латентного демографического поведения
- ретроанализ неполных данных
- опережающее статистику получение информации
- уточнение локации демографических событий