**ПРИЛОЖЕНИЕ А**

**Исходные данные исследования**

Таблица А.1. – Описание исходных переменных

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Показатель** | **Ед. изм.** | **Наименование** | **Тип переменной** | **Источник** |
| esales | млн. руб. | Оборот розничной интернет торговли, в фактически действовавших ценах | y | Росстат |
| citypop | % | Доля городского населения, от общей численности населения | соц х | Росстат |
| youthpop | % | Доля населения в трудоспособном возрасте, от общей численности населения | соц х | Росстат |
| poorpop | % | Численность населения с денежными доходами ниже границы бедности (величины прожиточного минимума), от общей численности населения | соц х | Росстат |
| incomepop | руб/мес. | Среднедушевые денежные доходы населения руб./мес. | соц х | Росстат |
| pop | человек | Численность постоянного населения в среднем за год | соц х | Росстат |
| students | человек | Численности студентов образовательных организаций, осуществляющих образовательную деятельность по образовательным программам высшего образования | соц х | Министерство образования |
| covid |  | Фиктивная переменная: до 2020 - 0, после 2020 - 1 | соц х |  |
| grp | млн. руб | Валовой региональный продукт в тек. ценах | эк х | Росстат |
| entrepreneur | человек | Численность индивидуальных предпринимателей в среднем за год, сведения о которых содержатся в Едином государственном реестре индивидуальных предпринимателей | эк х | ФНС |
| investment | % | Доля инвестиций, направленных на реконструкцию и модернизацию, в общем объеме инвестиций в основной капитал | эк х | Росстат |
| cpi | индекс | Индекс потребительских цен, к прошлому году | эк х | Росстат |
| card | шт. | Количество платежных карт, выданных физ лицам на территории региона на конец периода | тех х | Банк России |
| internetorg | % | Доля организаций, использующих широкополосный доступ к сети Интернет, в общем числе организаций (значение показателя за год) | тех х | Росстат |
| internetpop | % | Использование сети Интернет населением в возрасте 15 лет и старше, от общей численности населения в возрасте 15 лет и старше, соответствующего субъекта Российской Федерации | тех х | Росстат |

Продолжение таблицы А.1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| sctec | % | Доля продукции высокотехнологичных и наукоемких отраслей в валовом региональном продукте | тех х | Росстат |
| citydig | % | Уровень цифровизации местной телефонной сети в городской местности по субъектам Российской Федерации | тех х | Минцифры России |
| condig | % | Уровень цифровизации местной телефонной сети в сельской местности по субъектам Российской Федерации | тех х | Минцифры России |
| roads | км | Протяженность автомобильных дорог общего пользования федерального, регионального или межмуниципального и местного значения с твердым покрытием на конец года | лог х | Росстат |
| cargo | млн. тонно-км | Грузооборот автомобильного транспорта по субъектам Российской Федерации | лог х | Росстат |

**ПРИЛОЖЕНИЕ В**

**Промежуточные результаты расчетов и проверок на гипотезы**

Таблица В.1. – Анализ исследований в области прогнозирования продаж в электронной коммерции

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Наименование, год | Авторы | Набор данных | Основная модель | Модели для сравнения | Результаты |
| Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology, 2019 [16] | K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, B. Seaman | Данные о продажах с сайта Walmart.com, 190 дней за 2018 год | LSTM на разных уровнях разделения данных (по категориям, все вместе, по группам), прогноз на 10 дней продаж | EWMA, ETS, ARIMA, Prophet, Naïve Model | Наилучшее значение MAPE (0,871) показала модель LSTM, основанная на групповом разделении данных, которая показывает, что иерархия продуктов электронной коммерции содержат различные модели межпродуктового спроса и корреляции.  MAPE других моделей: ARIMA – 1,084; EWMA – 1,146; ETS – 1,097; Naïve Model – 0,981; Prophet – 1,087 |
| Sales Forecast in E-commerce using Convolutional Neural Network, 2017 [17] | K. Zhao, C. Wang | Данные о продажах CaiNiao.com, 1814892 записи о 1963 товаров в 5 регионах за период с 10.10.2014 по 27.12.2015, прогноз на 25 дней продаж | CNN, CNN+WD, CNN+WD+TL | ARIMA, FE+GBRT, DNN | Наилучшее значение MSE (126,22) имеет модель CNN+WD+TL (конволюционная нейронная сеть на структурированных данных с использованием метода трансферного обучения и метода снижения веса).  MSE других моделей: ARIMA – 175,16; FE+GBRT – 156,07; DNN – 156,48; CNN – 145,69; CNN+WD – 131,86 |

Продолжение таблицы В.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Analysis and Forecast of the Use of E-Commerce in Enterprises of the European Union States, 2022 [23] | G. Soava, A. Mehedintu, M. Sterpu | Данные о динамике доли предприятий ЕС, осуществляющих электронную коммерцию, и доли в общем обороте, полученной от электронной коммерции, на период 2003–2020 гг, прогноз на 2025 год. | Регрессионные модели | Модель авторегресии первого (1), второго (2) порядков, модель авторегрессии второго порядка с линейным трендом (3), модель линейной временной регрессии (4), квадратичная модель регрессии (5), кубическая модель регрессии (6) | Темпы роста продаж в электронной коммерции достигли зрелости в развитых странах.  Полиномиальная кубическая регрессия (6) статистически обеспечивает лучшую оценку ряда данных среди шести протестированных моделей (R2 – 0,.956; RMSE – 0,667).  R2, RMSE других моделей: (1) – 0,856; 1,159; (2) – 0,897; 0,822; (3) – 0,948; 0,778; (4) – 0,946; 0,733; (5) – 0,950; 0,704 |
| E-commerce Sales Forecast Based on Ensemble Learning, 2021 [18] | C. Zhan, J. Li, W. Jiang, W. Sha and Y. Guo | Данные о 75 000 000 транзакций компании электронной коммерции с 01.2013 – 07.2019 гг., прогноз на 7 дней продаж | Алгоритмы машинного обучения | GBDT, CatBoost, LightGBM, XGBoost | Наилучшие значения показывает алгоритм CatBoost (R2 – 0,602; RMSE – 17,24).  R2, RMSE других моделей: GBDT – 0,575; 17,932; LightGBM – 0,600; 17,282; XGBoost – 0,597; 17,349 |

Продолжение таблицы В.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise, 2019 [8] | S. Ji, X. Wang, W. Zhao, D. Guo | Данные о продажах, ассортименте, поведении покупателей и т.п. с сайта Jollychic за 2017 год, прогноз на месяц | CA-XGBoost (кластеризация+моделирование линейной части с помощью ARIMA, XGBoost – для моделирования нелинейеной части) | ARIMA, XGBoost, C-XGBoost, A-XGBoost | CA-XGBoost превосходит другие модели (RMSE – 3,282), метод обеспечил теоретическую поддержку прогноза продаж компании электронной коммерции и может служить эталоном. Компании электронной коммерции желательно выбирать разные модели прогнозирования для разных товаров вместо использования одной.  RMSE других моделей: ARIMA – 21,931; XGBoost – 6,049; A-XGBoost – 6,287; C-XGBoost – 4,832 |
| E-Commerce System for Sale Prediction Using Machine Learning Technique, 2020 [19] | Karandeep Singh, P M Booma, Umapathy Eaganathan | Данные об истории 100 000 заказов компании Olist Store (бразильский интернет-магазин) с 2016 по 2018 гг. | Random Forest, Gradient Boost | ARIMA, SARIMA | Точность обучения Gradient Boost показал значение выше, чем Random Forest (99,99% против 90,9%), при этом r2 выше у метода Random Forest (87% против 85%). Также Random Forest показал лучше результат по RMSE (76,669 против 82,230 у Gradient Boost). Также исследователи признали модель SARIMA лучшей в построении регрессии по показателю MAPE (6,69% в сравнении с 8.62% у ARIMA). |
| A multi-series framework for demand forecasts in E-commerce, 2019 [20] | R. Garnier, A. Belletoile | Данные сайта Cdiscount.com о продажах 99305 товаров в 1031 категории, прогноз на 6 недель продаж | XGBoost | ES, Random Forest | XGBoost с учетом сезонности показывает наилучшие результаты по RMSE – 0,725.  RMSE других моделей: ES – 1,09; Random Forest – 0,831; XGBoost без учета сезонности – 0,730 |

Продолжение таблицы В.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Short-Term Demand Forecast of E-Commerce Platform Based on ConvLSTM Network, 2022 [21] | Z. Li, N. Zhang | Данные о продажах компании Q: 5 магазинов, охватывающих период с 1 января 2017 года по 31 декабря 2018 года, в общей сложности 11 393 481 единица данных, включая около 500 видов товаров | ConvLSTM | LightGBM, LSTM | ConvLSTM показала лучшие результаты среди трех моделей прогнозирования, со средним значением MAPE 38,3, что на 0,68 процентных пункта ниже, чем у модели LGBM, и на 0,42 процентных пункта ниже, чем у модели LSTM |
| Prediction of the future trend of e-commerce, 2021 [24] | Engström F.,  Nilsson Rojas D. | Данные опроса пользователей в Швеции, проведенной Internetsif-telsen, Svenscarna och Internet за период с 2000 по 2020 года и с 2015 по 2020 года. | Регрессионные модели | Полиномиальная регрессия, SVR, ARIMA | ARIMA показала самые низкие наилучшие значения MSE – 4, 4926.  MSE других моделей: Полиномиальная регрессия – 173, 031; SVR – 721,046 |
| Использование методов машинного обучения для решения задач прогнозирования спроса на новый товар в интернет-маркетплейсе, 2020 [22] | Осин А.А., Фомин А.К., Сологуб Г.Б, Виноградов В.И. | Данные об истории продаж и хранении товаров на сайте Ozon.ru: 1886147 уникальных продаж, история товаров на складе размером 13632607 уникальных значений | LightGBM | - | Построенная модель LightGBM показала следующие значения: R2 = 0,0367 и MSE = 330,086 |

Продолжение таблицы В.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Применение методов машинного обучения в области прогнозирования объема продаж с учетом динамически изменяющихся признаков, 2020 [9] | Валиахметова Ю.И., Идрисова Э.И. | Данные о продажах компании Walmart за 5,5 лет с 2011 по 2016 год, прогноз на меясц | CatBoost | ARIMA, SARIMA, Аддитивная регрессионная модель, LSTM | Регрессионные модели показывают хорошие результаты на сгруппированных временных рядах. Хорошие оценки показала модель SARIMA. Но в ходе испытаний этих методов на сгруппированных и отдельных временных рядах алгоритм CatBoost показал наилучшие результаты (R2 – 0,976, RMSE – 0,664).  R2, RMSE других моделей: SARIMAX – 0,865; 2605,7811; LSTM – 0,599; 3,039 |

Таблица В.2. – Описательная статистика для всех данных

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Показатель** | **Кол-во наблюдений** | **Среднее значение** | **Стандартное отклонение** | **Минимальное значение** | **Квартили распределения** | | | | | **Максимальное значение** |
| **count** | **mean** | **std** | **min** | **10%** | **30%** | **50%** | **70%** | **90%** | **max** |
| **esales** | 486 | 10814,42 | 41757,65 | 0 | 125,74 | 618,17 | 1679,27 | 4783,19 | 16972,56 | 555569,3 |
| **eshare** | 486 | 1,38 | 1,66 | 0 | 0,1 | 0,4 | 0,8 | 1,5 | 3,4 | 10,4 |
| **citypop** | 486 | 70,39 | 13,09 | 29 | 53,1 | 66,25 | 71,9 | 77,72 | 82,52 | 100 |
| **youthpop** | 486 | 75,22 | 4,26 | 68,7 | 70,9 | 72,5 | 74,6 | 76,6 | 80,1 | 90,4 |
| **poorpop** | 486 | 14,52 | 5,04 | 5 | 8,85 | 12,2 | 13,8 | 15,7 | 20,8 | 37,8 |
| **incomepop** | 486 | 29654,6 | 10182,19 | 14963 | 20631,5 | 24156,5 | 27147 | 30610,5 | 41847,5 | 87972 |
| **pop** | 486 | 1808925 | 1820073 | 138401 | 465844 | 865581,5 | 1199746 | 1942652 | 3779787 | 12666565 |
| **students** | 486 | 51081,44 | 86937,63 | 1947 | 7866,5 | 18629,5 | 29746,5 | 49952,5 | 97000,5 | 759324 |
| **covid** | 486 | 0,33 | 0,47 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| **grp** | 486 | 1140211 | 2413590 | 47434,9 | 153425,3 | 315598,5 | 539976,7 | 881439,4 | 2008523 | 24471160 |
| **entrepreneur** | 486 | 44736,52 | 51276,79 | 2434 | 11587 | 18578,5 | 29846,5 | 45752,5 | 82152 | 353944 |
| **investment** | 486 | 18,85 | 7,79 | 3,5 | 8,9 | 14,6 | 18,05 | 22,45 | 30,15 | 45 |
| **cpi** | 486 | 104,58 | 2,02 | 100,84 | 102,39 | 103,22 | 104,01 | 105,78 | 107,6 | 111,88 |
| **card** | 486 | 3469287 | 6413238 | 184566 | 539955,5 | 1322701 | 1848169 | 3225757 | 6667488 | 64926530 |
| **internetorg** | 486 | 78,72 | 11,4 | 29 | 61,2 | 75,05 | 81,7 | 86,35 | 90,15 | 97,7 |
| **internetpop** | 486 | 81,11 | 5,86 | 65,3 | 73,75 | 77,6 | 80,85 | 83,9 | 89,2 | 97,4 |
| **sctec** | 486 | 19,32 | 5,53 | 6,7 | 11,95 | 16,7 | 19 | 21,7 | 27,15 | 34,7 |
| **citydig** | 486 | 93,86 | 8,49 | 44,7 | 82,55 | 92,6 | 97,3 | 99,7 | 100 | 100 |
| **condig** | 486 | 85,66 | 16,51 | 22,4 | 61,5 | 79,2 | 91,85 | 99,6 | 100 | 100 |
| **roads** | 486 | 13340,66 | 8752,6 | 855,86 | 3539,36 | 7952,8 | 10943,01 | 16848,34 | 24440,81 | 45687,84 |
| **cargo** | 486 | 2253,19 | 2561,51 | 9,9 | 185,46 | 733,3 | 1506,82 | 2377,82 | 5574,3 | 22632 |

Таблица В.3. – Описательная статистика для данных за 2016-2020 гг.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Показатель** | **Кол-во наблюдений** | **Среднее значение** | **Стандартное отклонение** | **Минимальное значение** | **Квартили распределения** | | | | | **Максимальное значение** |
| **count** | **mean** | **std** | **min** | **10%** | **30%** | **50%** | **70%** | **90%** | **max** |
| **esales** | 405 | 8166,65 | 32404,81 | 0 | 122,26 | 531,97 | 1367,35 | 3431,46 | 11003,69 | 481413,77 |
| **eshare** | 405 | 1,03 | 1,26 | 0 | 0,1 | 0,3 | 0,7 | 1,1 | 2,36 | 9,3 |
| **citypop** | 405 | 70,32 | 13,12 | 29 | 52,92 | 66,11 | 71,8 | 77,69 | 82,38 | 100 |
| **youthpop** | 405 | 75,25 | 4,28 | 68,7 | 70,9 | 72,5 | 74,6 | 76,6 | 80,26 | 90,4 |
| **poorpop** | 405 | 14,76 | 5,09 | 5,9 | 9 | 12,5 | 13,9 | 15,88 | 21,46 | 37,8 |
| **incomepop** | 405 | 28808,61 | 9627,02 | 14963 | 20246,8 | 23846,4 | 26483 | 29769,2 | 41386,4 | 78106 |
| **pop** | 405 | 1810673,86 | 1817552,71 | 139592 | 465701,2 | 865316,6 | 1205584 | 1951684,4 | 3756477,8 | 12666565 |
| **students** | 405 | 51361,74 | 86182,73 | 1947 | 7943,6 | 18706,2 | 29938 | 50400,2 | 97144,6 | 736274 |
| **covid** | 405 | 0,2 | 0,4 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| **grp** | 405 | 1069371,93 | 2237835,11 | 47434,9 | 150296,62 | 299540,22 | 521051,5 | 849588,98 | 1777738,18 | 20260717,7 |
| **entrepreneur** | 405 | 44997,25 | 50556,76 | 2887 | 11885,6 | 18694 | 30377 | 46213,4 | 82429,8 | 348621 |
| **investment** | 405 | 18,88 | 7,81 | 3,5 | 8,94 | 14,52 | 18,1 | 22,7 | 30,2 | 43,3 |
| **cpi** | 405 | 104,13 | 1,88 | 100,84 | 102,26 | 102,98 | 103,65 | 104,45 | 107,36 | 111,88 |
| **card** | 405 | 3350743,26 | 6159827,64 | 184566 | 521384,6 | 1286524,6 | 1796344 | 3189716,6 | 6468084,6 | 56830080 |
| **internetorg** | 405 | 79,17 | 12,22 | 29 | 59,98 | 76,28 | 83,8 | 87,18 | 91,16 | 97,7 |
| **internetpop** | 405 | 80,13 | 5,61 | 65,3 | 73,34 | 77,2 | 79,7 | 82,8 | 88,26 | 94,8 |
| **sctec** | 405 | 19,3 | 5,39 | 6,7 | 12,04 | 16,72 | 19 | 21,6 | 26,66 | 34,7 |
| **citydig** | 405 | 93,39 | 8,77 | 44,7 | 81,8 | 91,9 | 96,7 | 99,6 | 100 | 100 |
| **condig** | 405 | 84,75 | 16,95 | 22,4 | 60,72 | 78,57 | 91,2 | 99 | 100 | 100 |
| **roads** | 405 | 13276,77 | 8717,39 | 855,86 | 3538,71 | 7916 | 10871,5 | 16826,22 | 24361,01 | 44883,87 |
| **cargo** | 405 | 2173,44 | 2408,61 | 9,9 | 200,45 | 722,35 | 1497,55 | 2376,92 | 5321,24 | 22632 |

Таблица В.4. – Описательная статистика для данных за 2021 год

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Показатель** | **Кол-во наблюдений** | **Среднее значение** | **Стандартное отклонение** | **Минимальное значение** | **Квартили распределения** | | | | | **Максимальное значение** |
| **count** | **mean** | **std** | **min** | **10%** | **30%** | **50%** | **70%** | **90%** | **max** |
| **esales** | 81 | 24759,04 | 71541,7 | 89,84 | 657,57 | 4256,43 | 6511,14 | 13434,33 | 46675,73 | 555569,3 |
| **eshare** | 81 | 3,11 | 2,24 | 0,1 | 0,8 | 1,7 | 2,4 | 3,8 | 5,6 | 10,4 |
| **citypop** | 81 | 70,7 | 13,04 | 29,1 | 54,9 | 66,7 | 72,2 | 77,9 | 82,7 | 100 |
| **youthpop** | 81 | 75,1 | 4,16 | 69,3 | 70,9 | 72,4 | 74,7 | 76,7 | 79,8 | 89,5 |
| **poorpop** | 81 | 13,35 | 4,64 | 5 | 7,9 | 11,4 | 12,8 | 14,4 | 19,3 | 29,3 |
| **incomepop** | 81 | 33884,56 | 11780,08 | 17954 | 23773 | 27788 | 30505 | 34918 | 49938 | 87972 |
| **pop** | 81 | 1800179 | 1843984 | 138401 | 484898 | 869839 | 1175726 | 1902148 | 3792279 | 12645258 |
| **students** | 81 | 49679,94 | 91160,37 | 1957 | 7571 | 18194 | 27747 | 46567 | 95712 | 759324 |
| **grp** | 81 | 1494405 | 3142331 | 71336,2 | 199326,3 | 421672,8 | 690253,2 | 1083779 | 2122537 | 24471160 |
| **entrepreneur** | 81 | 43432,89 | 55048,99 | 2434 | 10924 | 17032 | 27559 | 41303 | 77718 | 353944 |
| **investment** | 81 | 18,66 | 7,77 | 4,4 | 8,9 | 14,7 | 18 | 21,2 | 28 | 45 |
| **cpi** | 81 | 106,79 | 1 | 104,32 | 105,49 | 106,35 | 106,86 | 107,41 | 107,93 | 108,96 |
| **card** | 81 | 4062006 | 7570133 | 259194 | 692187 | 1644919 | 2236208 | 3825353 | 7730676 | 64926530 |
| **internetorg** | 81 | 76,47 | 5,2 | 57,5 | 70,6 | 73,9 | 76 | 79,4 | 82,2 | 87,6 |
| **internetpop** | 81 | 85,97 | 4,54 | 76,3 | 80,6 | 83,1 | 85,1 | 88 | 92,2 | 97,4 |
| **sctec** | 81 | 19,42 | 6,21 | 8,2 | 10,4 | 16,2 | 18,9 | 22 | 29,2 | 33,5 |
| **citydig** | 81 | 96,2 | 6,49 | 66,2 | 87,2 | 96,9 | 99,2 | 100 | 100 | 100 |
| **condig** | 81 | 90,17 | 13,33 | 42 | 73,4 | 88,4 | 96,5 | 100 | 100 | 100 |
| **roads** | 81 | 13660,11 | 8974,87 | 1071,11 | 3654,14 | 8271,53 | 11100,82 | 17203,09 | 24838,99 | 45687,84 |
| **cargo** | 81 | 2651,94 | 3207,46 | 21,7 | 114,29 | 778,52 | 1597,63 | 2381,16 | 6873,17 | 14074,42 |

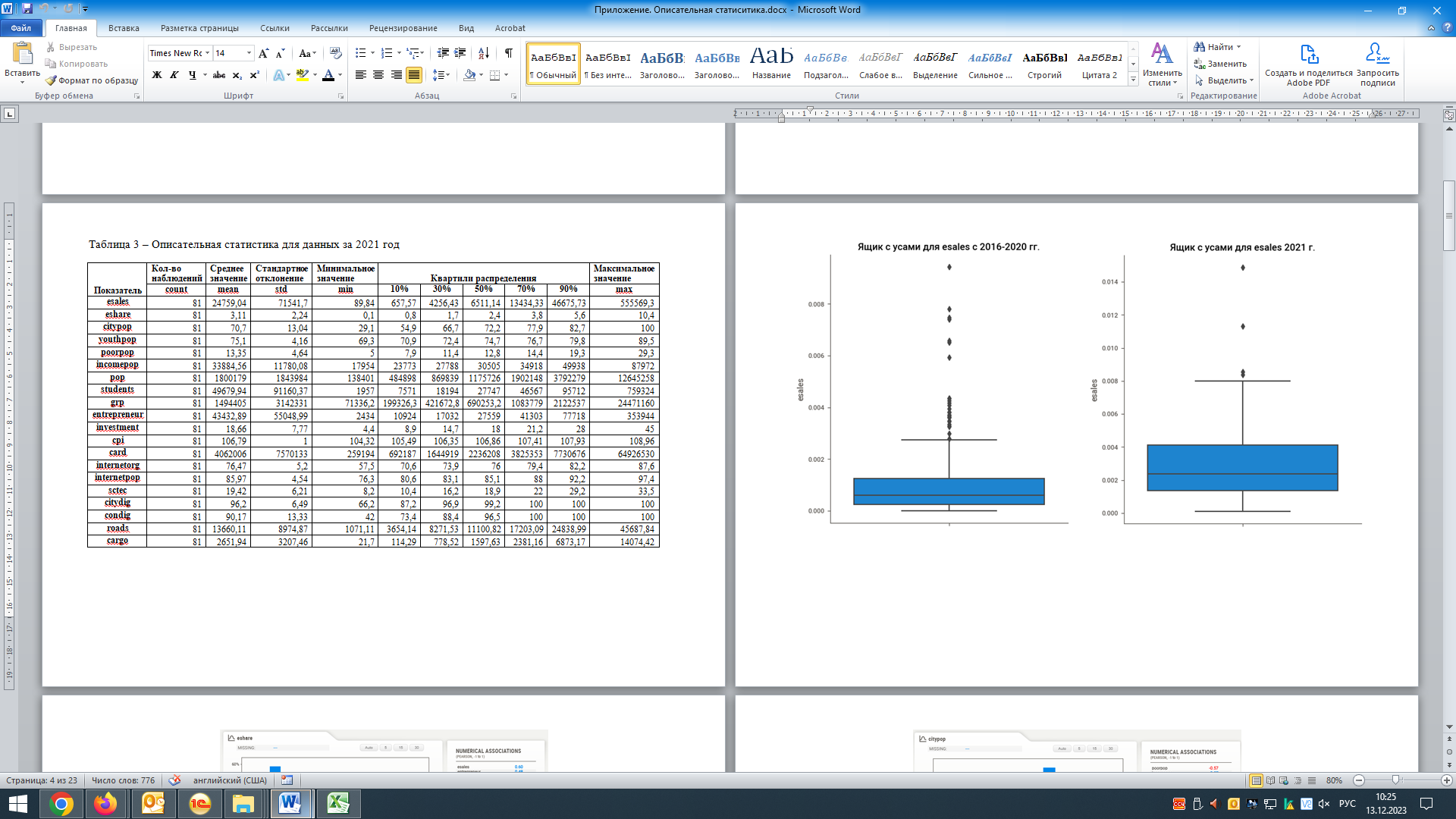


Рисунок В.1. – График «ящик с усам» для нормализованных данных по показателю esales   
по обучающей (за 2016-2020 гг.) и тестовой выборках (за 2021 г.)

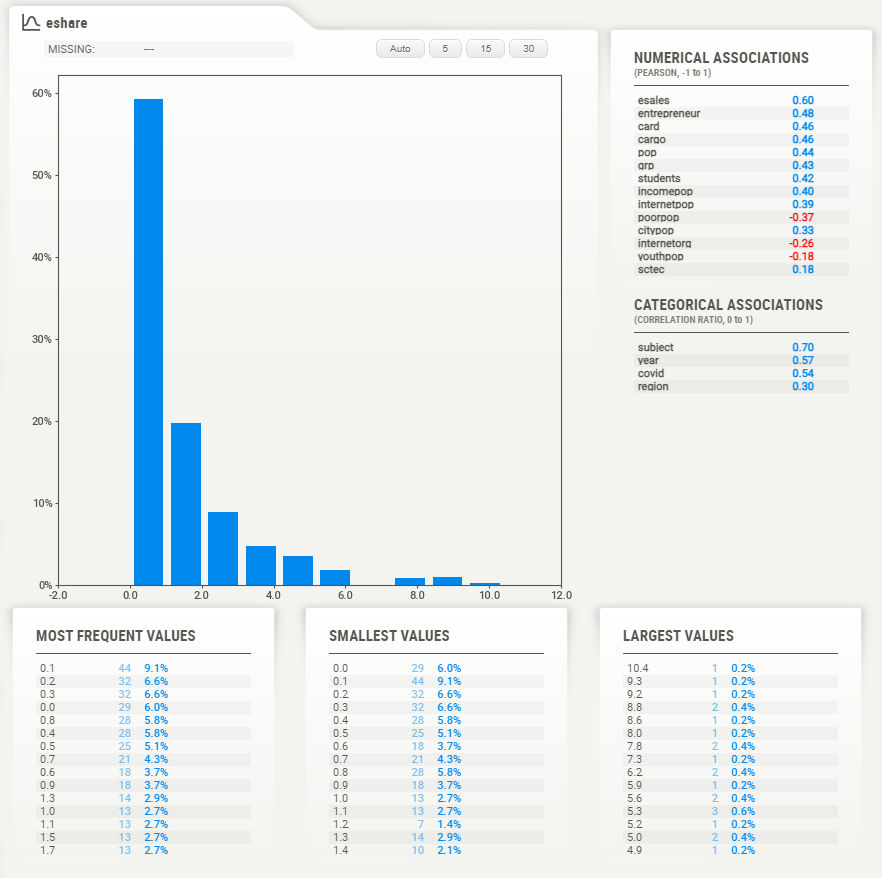


Рисунок В.2. – График распределения по показателю eshare

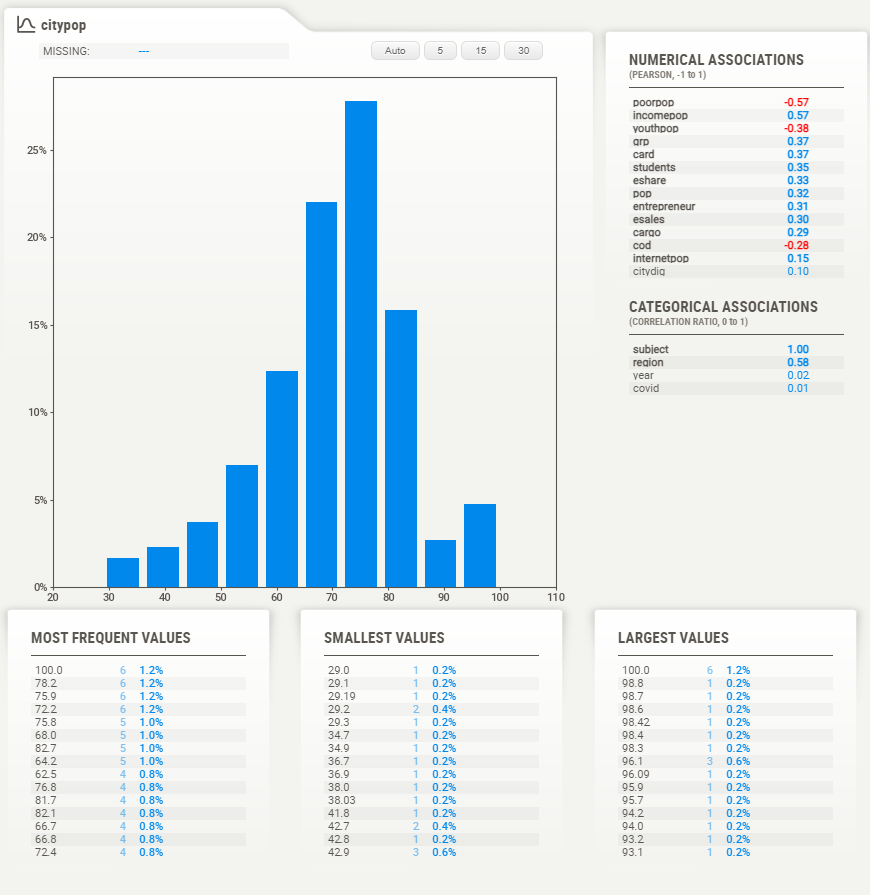


Рисунок В.3. – График распределения по показателю citypop

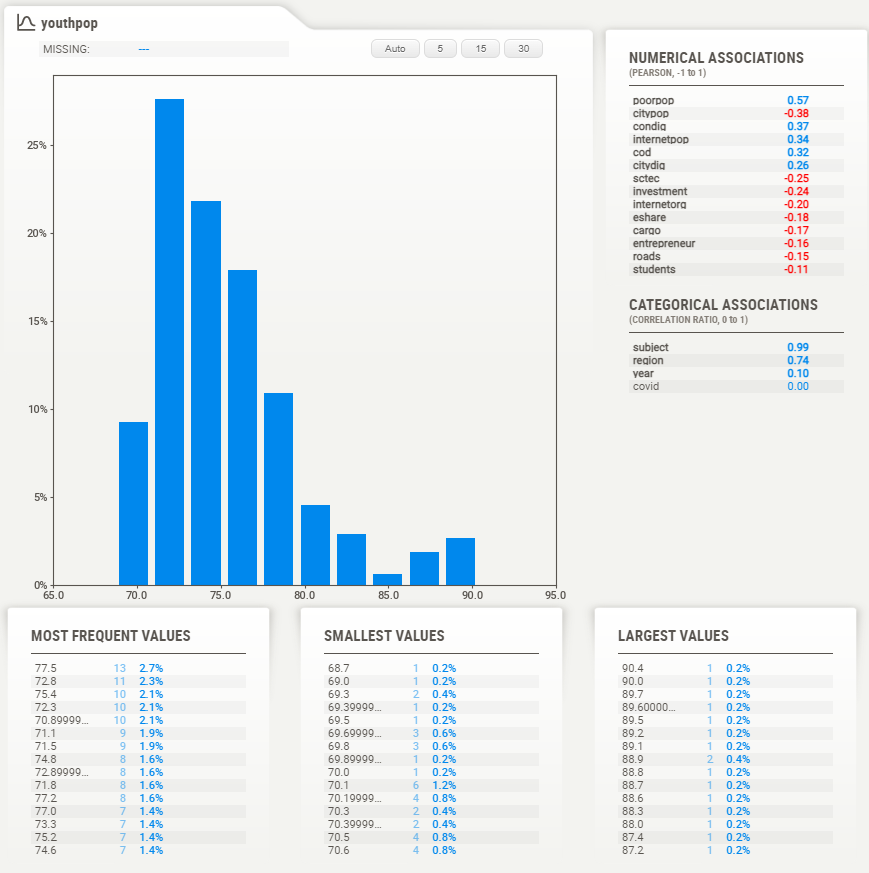


Рисунок В.4. – График распределения по показателю youthpop

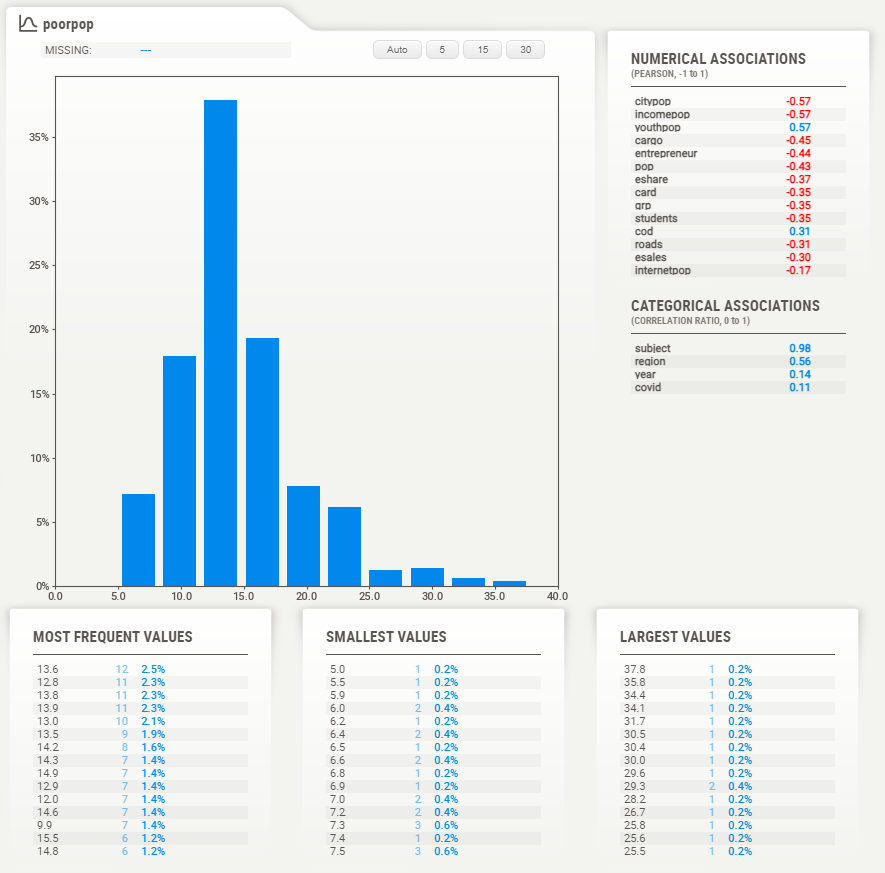


Рисунок В.5. – График распределения по показателю poorpop

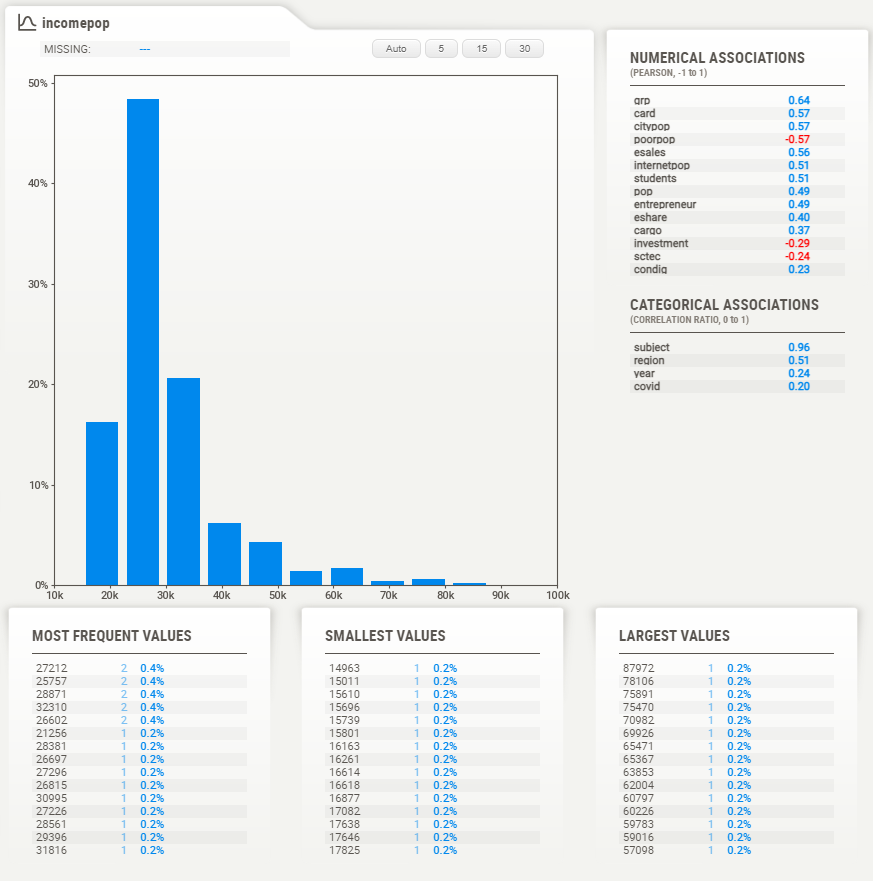


Рисунок В.6. – График распределения по показателю incomepop

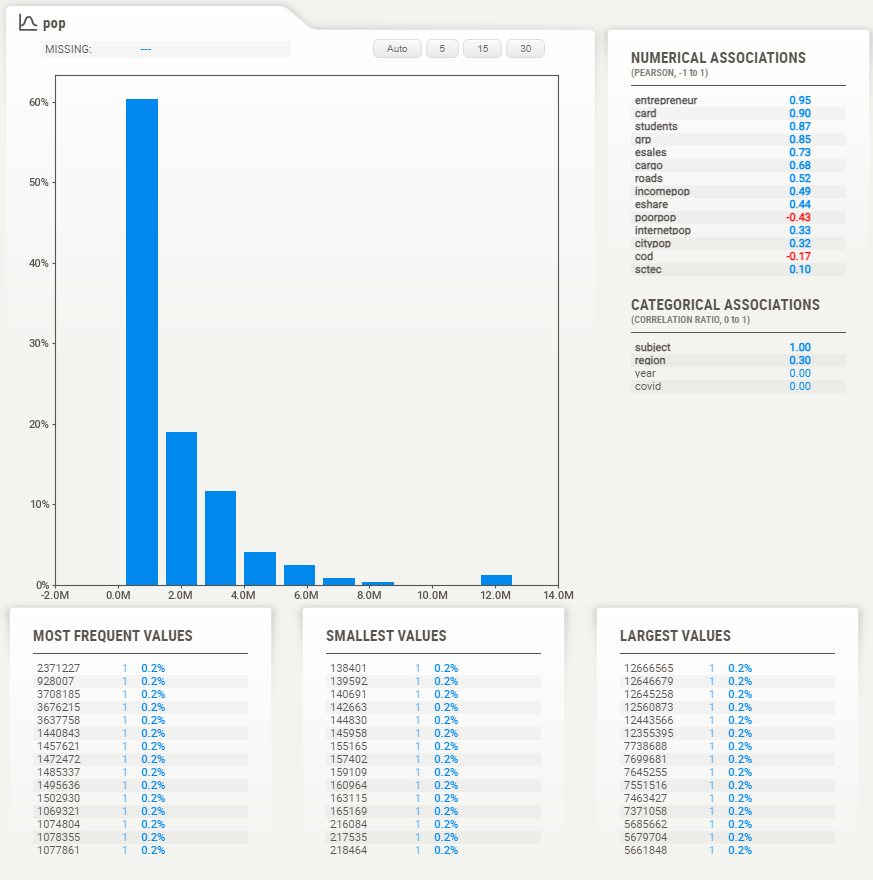


Рисунок В.7. – График распределения по показателю pop

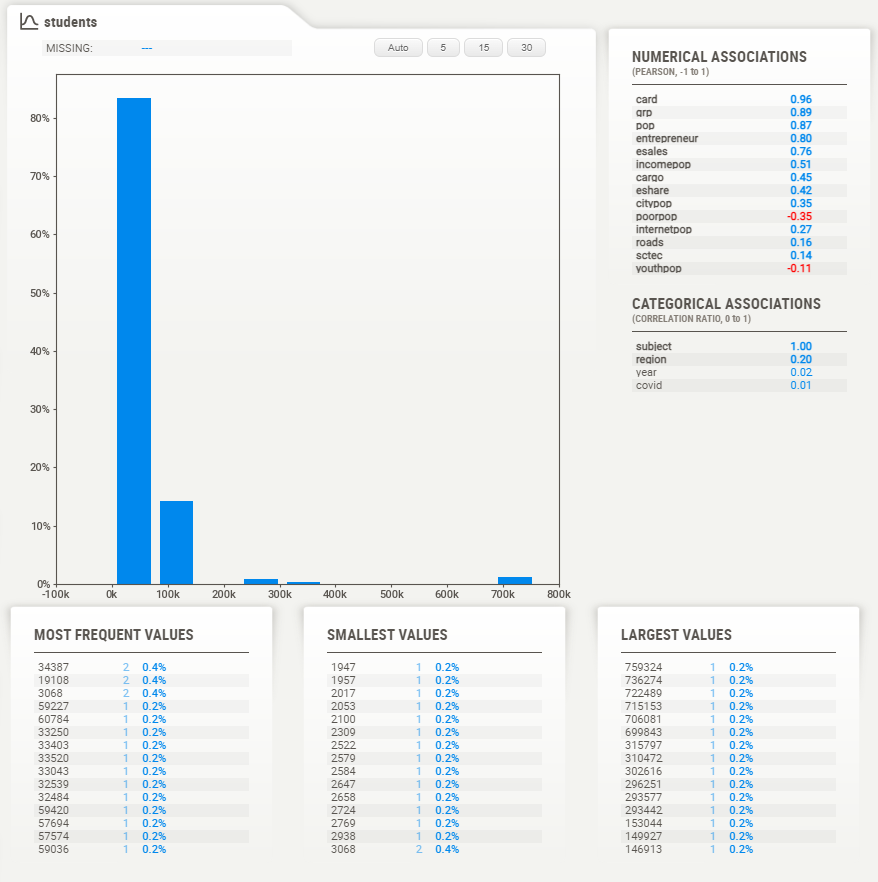


Рисунок В.8. – График распределения по показателю students

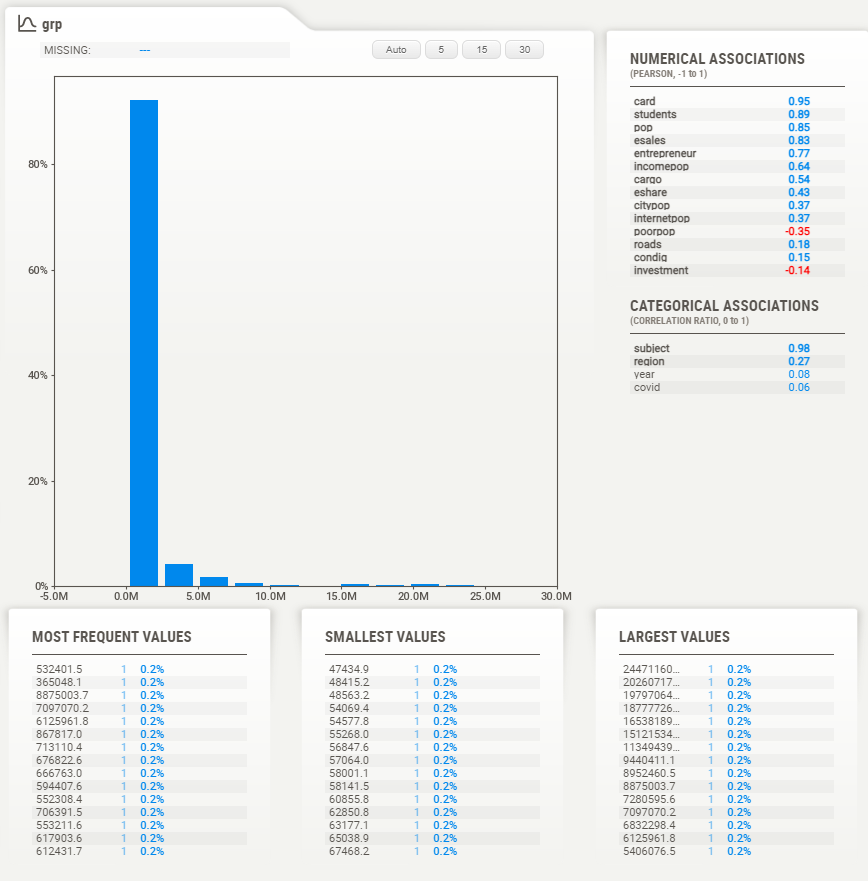


Рисунок В.9. – График распределения по показателю grp

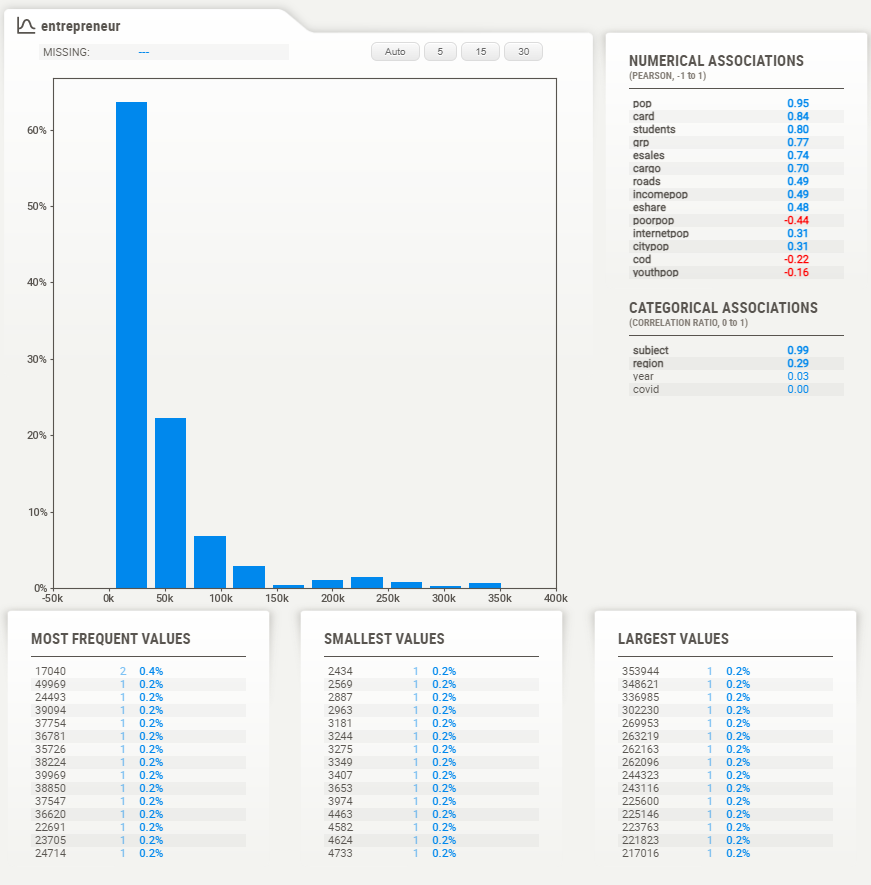


Рисунок В.10. – График распределения по показателю entrpreneur

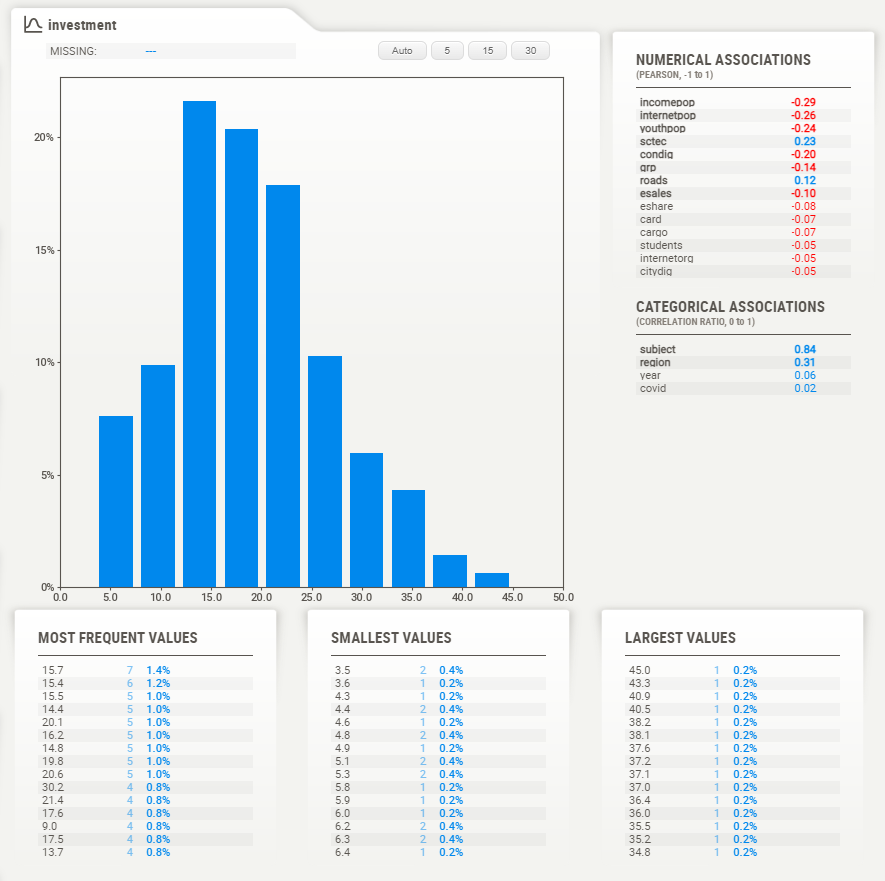


Рисунок В.11. – График распределения по показателю investment

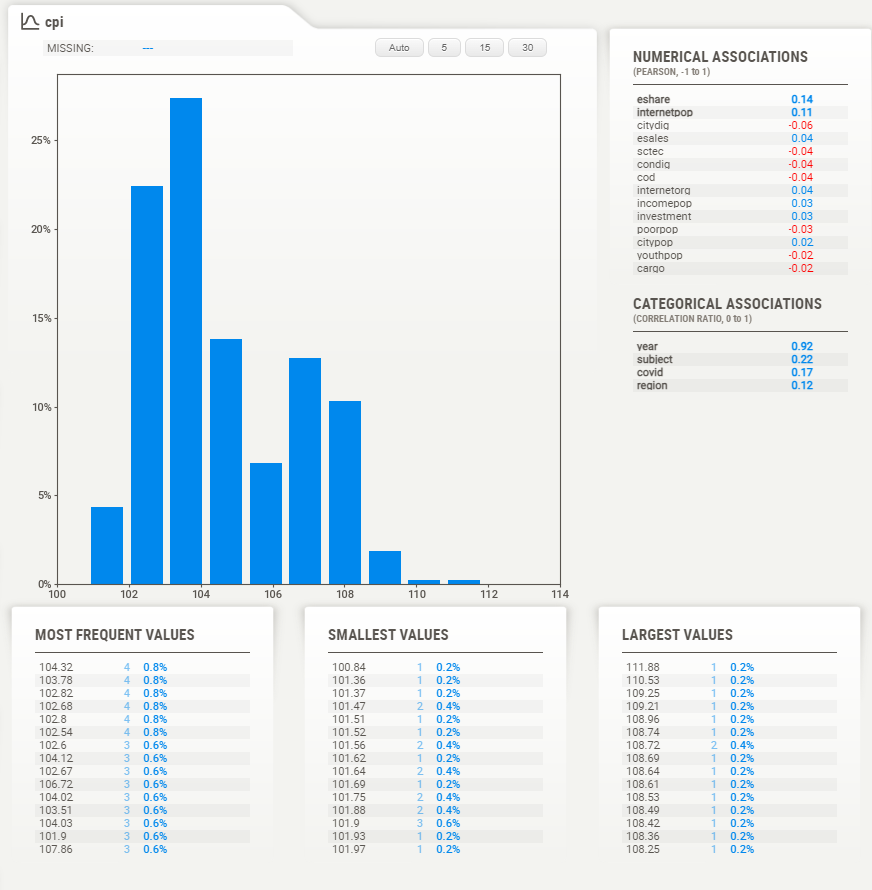


Рисунок В.12. – График распределения по показателю cpi

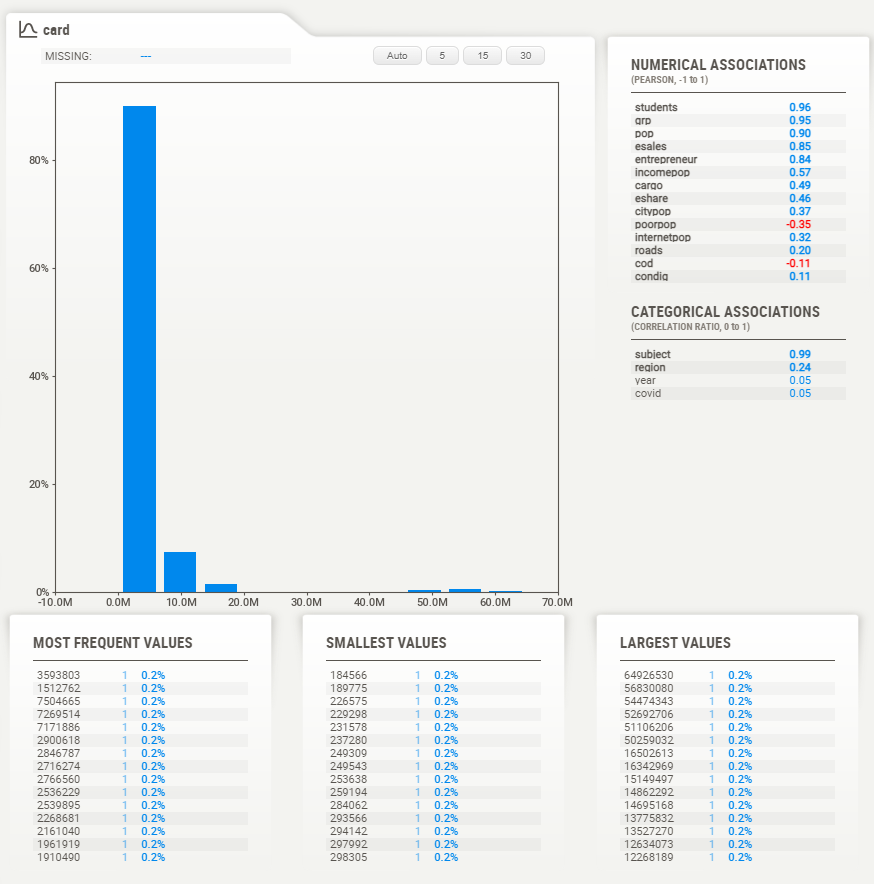


Рисунок В.13. – График распределения по показателю card

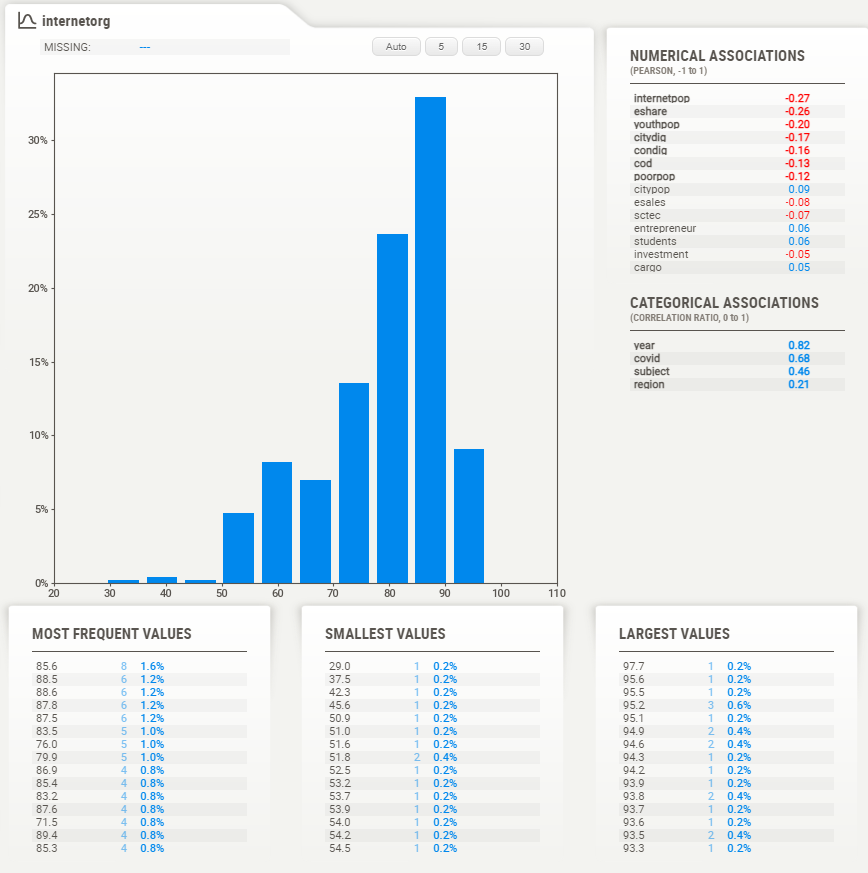


Рисунок В.14. – График распределения по показателю internetorg

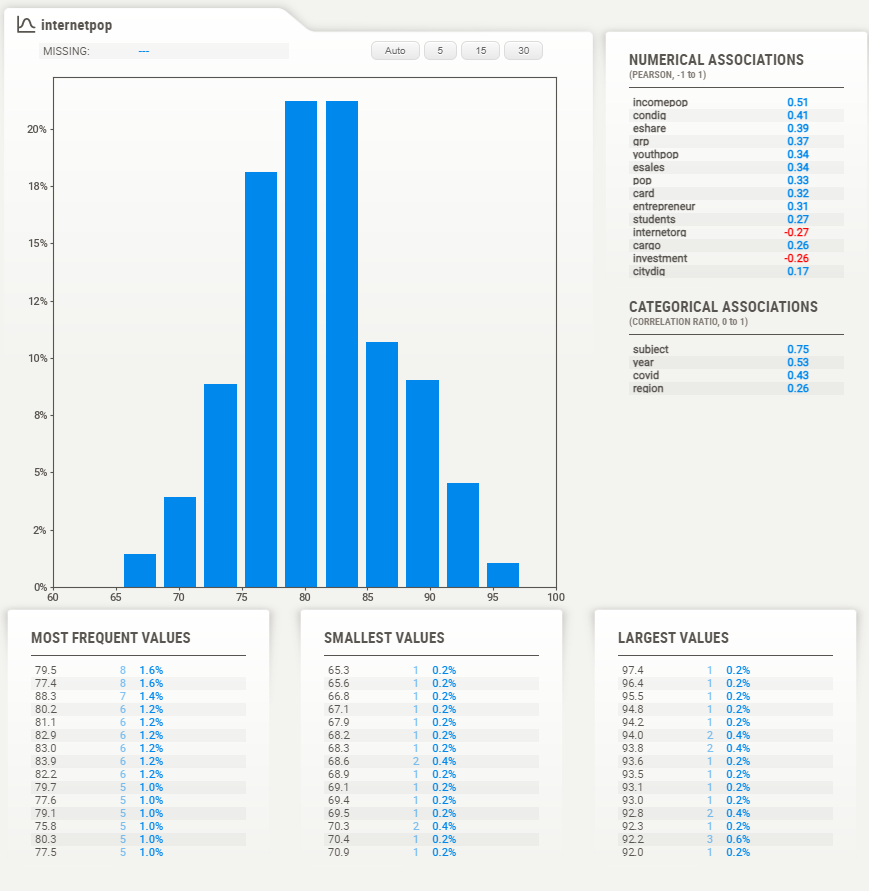


Рисунок В.15. – График распределения по показателю internetpop

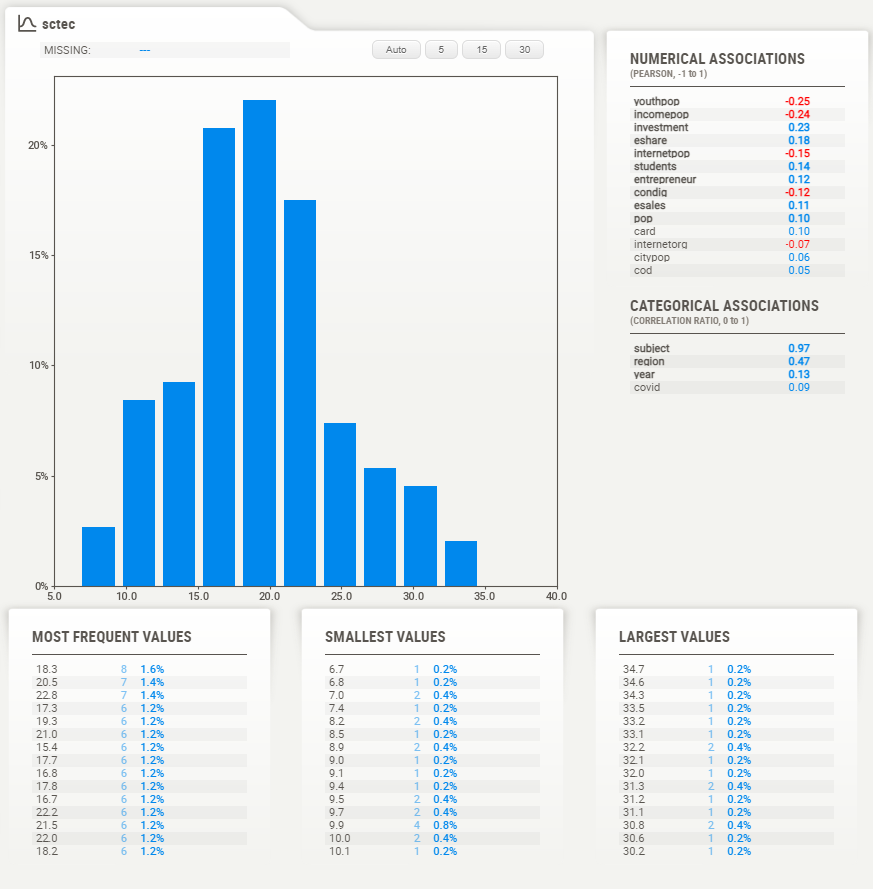


Рисунок В.16. – График распределения по показателю sctec

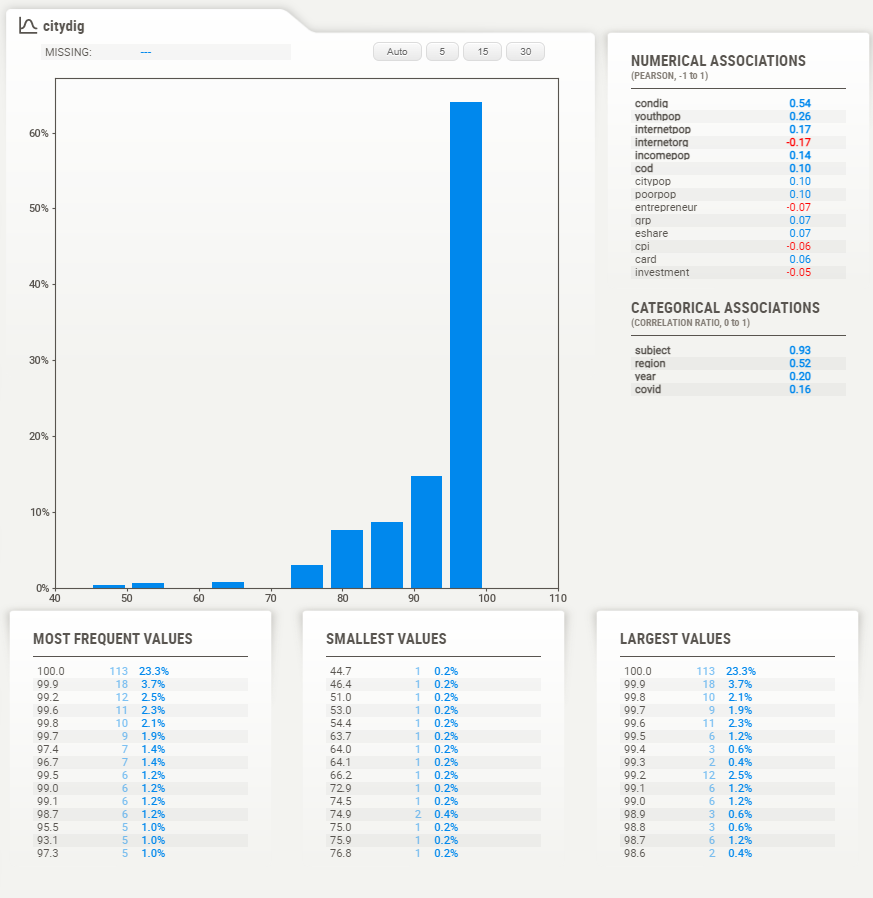


Рисунок В.17. – График распределения по показателю citydig

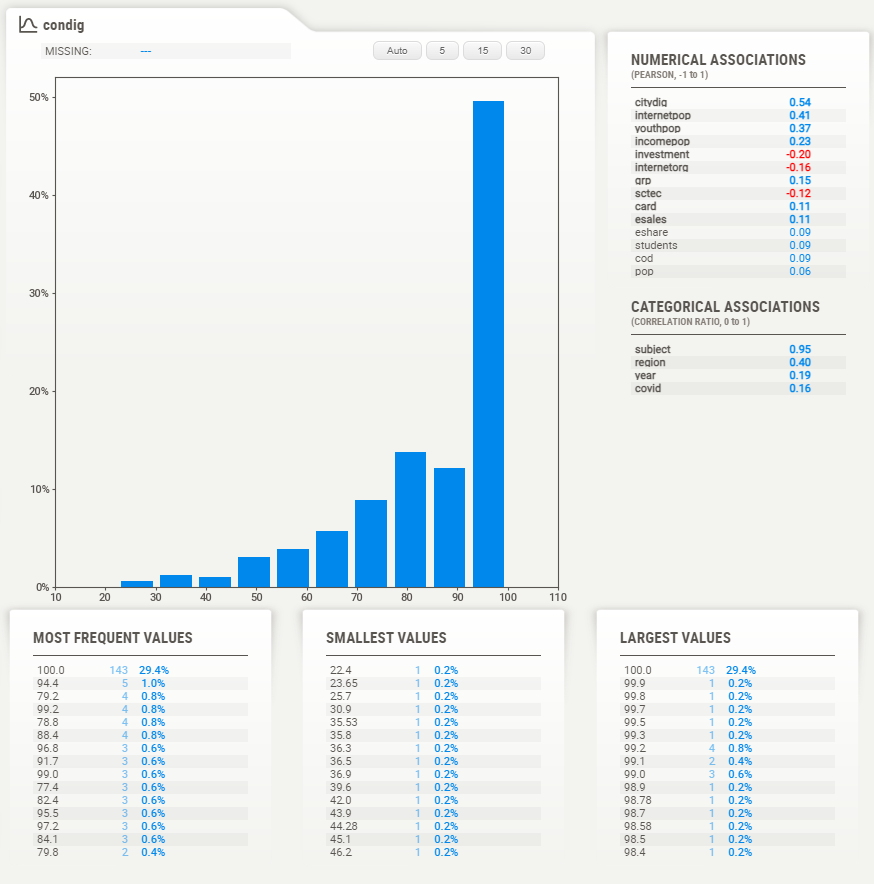


Рисунок В.18. – График распределения по показателю condig

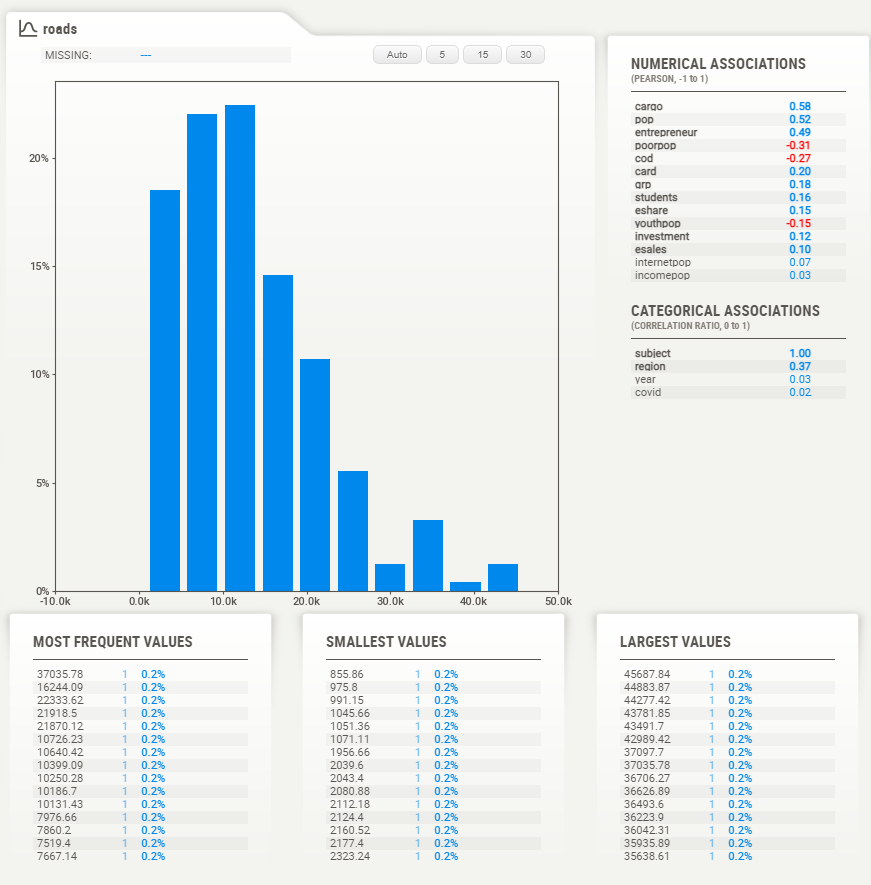


Рисунок В.19. – График распределения по показателю roads

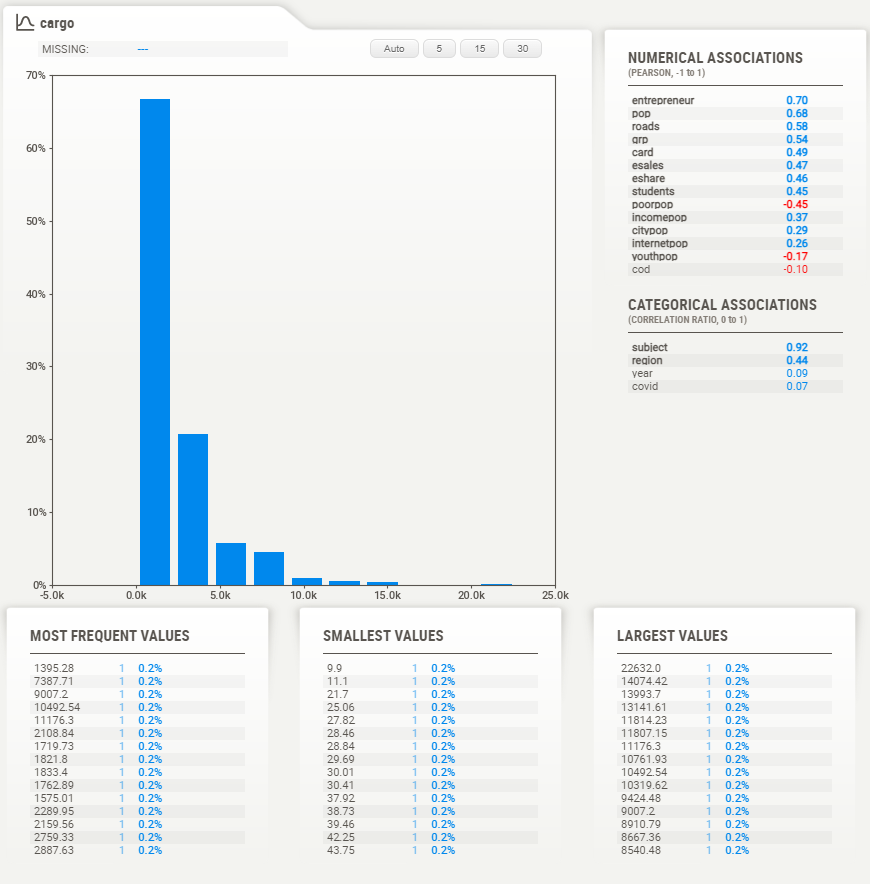


Рисунок В.20 – График распределения по показателю cargo

Таблица В.5. – Результаты пробной модели №0

|  |  |  |
| --- | --- | --- |
| **№** | **model0** | |
| **Выборка** | **train** | **test** |
| **Параметры модели** | iteration: 195; depth: 3; learning\_rate: 0.1 | |
| **Оценки** | | |
| **MAE** | 1179.4214 | 11633.2273 |
| **MAPE** | 2.7269 | 2.1668 |
| **SMAPE** | 70.2888 | 86.0580 |
| **RMSE** | 3 354 507.5064 | 607 991 750.8043 |
| **R2** | 0.9967 | 0.8797 |
| **CV-score (RMSE)** | 141.481 | |
| **Сумма esales за 2021 г, млн руб., факт** | 2 005 482.1600 | |
| **Сумма esales за 2021 г, млн руб., прогноз** | 1 196 716.9625 | |
| **Ошибка прогноза, млн руб.** | -808 765.1975 | |
| **Важность признаков** | | |
| **condig** | 0.109628 | |
| **youthpop** | 0.272583 | |
| **covid** | 0.400967 | |
| **citydig** | 0.688815 | |
| **investment** | 0.824427 | |
| **cpi** | 1.075246 | |
| **citypop** | 1.296429 | |
| **cargo** | 2.069836 | |
| **internetorg** | 2.164315 | |
| **sctec** | 2.527187 | |
| **roads** | 2.550372 | |
| **poorpop** | 2.679711 | |
| **internetpop** | 4.039878 | |
| **incomepop** | 6.718212 | |
| **grp** | 6.920873 | |
| **students** | 10.744928 | |
| **entrepreneur** | 23.657444 | |
| **card** | 31.259149 | |

Таблица В.6. – Результаты модели №1 – после удаления переменных

|  |  |  |
| --- | --- | --- |
| **№** | **model1** | |
| **Выборка** | **train** | **test** |
| **Параметры модели** | iteration: 309; depth: 3; learning\_rate: 0.1 | |
| **Оценки** | | |
| **MAE** | 1061.2162 | 8854.7652 |
| **MAPE** | 1.4071 | 1.3406 |
| **SMAPE** | 68.7612 | 57.9787 |
| **RMSE** | 2 844 240.9956 | 380 504 575.1329 |
| **R2** | 0.9972 | 0.8797 |
| **CV-score (RMSE)** | 136.864 | |
| **Сумма esales за 2021 г, млн руб., факт** | 2 005 482.1600 | |
| **Сумма esales за 2021 г, млн руб., прогноз** | 1 393 136.9319 | |
| **Ошибка прогноза, млн руб.** | -612 345.2280 | |
| **Важность признаков** | | |
| **condig** | - | |
| **youthpop** | - | |
| **covid** | 5.345030 | |
| **citydig** | - | |
| **investment** | - | |
| **cpi** | - | |
| **citypop** | - | |
| **cargo** | - | |
| **internetorg** | - | |
| **sctec** | - | |
| **roads** | - | |
| **poorpop** | 9.682582 | |
| **internetpop** | 6.857714 | |
| **incomepop** | - | |
| **grp** | 9.366020 | |
| **students** | 11.402242 | |
| **entrepreneur** | 27.937793 | |
| **card** | 29.408619 | |

Таблица В.7. – Результаты модели №2 по гиперпараметрам, найденным по CatBoost grid\_search

|  |  |  |
| --- | --- | --- |
| **№** | **model2** | |
| **Выборка** | **train** | **test** |
| **Параметры модели** | iteration: 54; depth: 7; learning\_rate: 0.5 | |
| **Оценки** | | |
| **MAE** | 585.8895 | 8703.7902 |
| **MAPE** | 7.6486 | 0.7516 |
| **SMAPE** | 55.7586 | 63.0250 |
| **RMSE** | 650 536.7041 | 359 861 892.8238 |
| **R2** | 0.9993 | 0.9288 |
| **CV-score (RMSE)** | 136.688 | |
| **Сумма esales за 2021 г, млн руб., факт** | 2 005 482.1600 | |
| **Сумма esales за 2021 г, млн руб., прогноз** | 1 402 796.9598 | |
| **Ошибка прогноза, млн руб.** | -602 685.2002 | |
| **Важность признаков** | | |
| **condig** | - | |
| **youthpop** | - | |
| **covid** | 14.887477 | |
| **citydig** | - | |
| **investment** | - | |
| **cpi** | - | |
| **citypop** | - | |
| **cargo** | - | |
| **internetorg** | - | |
| **sctec** | - | |
| **roads** | - | |
| **poorpop** | 6.663156 | |
| **internetpop** | 10.819155 | |
| **incomepop** | - | |
| **grp** | 21.337247 | |
| **students** | 6.754762 | |
| **entrepreneur** | 19.747615 | |
| **card** | 19.790588 | |

Таблица В.8. – Результаты модели №3 по гиперпараметрам, найденным по sklearn GridsearchCV

|  |  |  |
| --- | --- | --- |
| **№** | **model3** | |
| **Выборка** | **train** | **test** |
| **Параметры модели** | iteration: 3179; depth: 11; learning\_rate: 0.01 | |
| **Оценки** | | |
| **MAE** | 197.2111 | 9031.0369 |
| **MAPE** | 3.0779 | 0.7392 |
| **SMAPE** | 34.3293 | 51.0931 |
| **RMSE** | 66 460.9105 | 471 801 545.9537 |
| **R2** | 0.9999 | 0.9067 |
| **CV-score (RMSE)** | 130.023 | |
| **Сумма esales за 2021 г, млн руб., факт** | 2 005 482.1600 | |
| **Сумма esales за 2021 г, млн руб., прогноз** | 1 350 346.3461 | |
| **Ошибка прогноза, млн руб.** | -655 135.8139 | |
| **Важность признаков** | | |
| **condig** | - | |
| **youthpop** | - | |
| **covid** | 12.410426 | |
| **citydig** | - | |
| **investment** | - | |
| **cpi** | - | |
| **citypop** | - | |
| **cargo** | - | |
| **internetorg** | - | |
| **sctec** | - | |
| **roads** | - | |
| **poorpop** | 7.232354 | |
| **internetpop** | 11.133892 | |
| **incomepop** | - | |
| **grp** | 9.678400 | |
| **students** | 12.231069 | |
| **entrepreneur** | 26.754248 | |
| **card** | 20.559610 | |

Таблица В.9. – Результаты модели №4 – CatBoost+ридж-регрессия

|  |  |  |
| --- | --- | --- |
| **№** | **model4** | |
| **Выборка** | **train** | **test** |
| **Параметры модели** | iteration: 3179  depth: 7  learning\_rate: 0.5  alpha (ridge): 1 | |
| **Оценки** | | |
| **MAE** | 5497.9049 | 6837.5675 |
| **MAPE** | 1.4232 | 0.9622 |
| **SMAPE** | 125.6982 | 60.9385 |
| **RMSE** | 162 664 697.9054 | 147 716 906.8992 |
| **R2** | 0.8421 | 0.9708 |
| **CV-score (RMSE)** | 170.618 | |
| **Сумма esales за 2021 г, млн руб., факт** | 2 005 482.1600 | |
| **Сумма esales за 2021 г, млн руб., прогноз** | 1 964 750.6646 | |
| **Ошибка прогноза, млн руб.** | -40 731.4954 | |