

УДК 004.942+656.02

Кулакова Ирина Михайловна,

к.т.н., доцент кафедры «Вычислительные машины и комплексы»
ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: iyelkina@mail.ru

Лебедева Ольга Анатольевна,

к.т.н., доцент кафедры «Управление на автомобильном транспорте»,
ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: kravhome@mail.ru

Ерофеев Ермак Владимирович,

обучающийся гр. ИТБ-23-1,
ФГБОУ ВО «Ангарский государственный технический университет»,
e-mail: ermak.080@mail.ru

ОБРАБОТКА И АНАЛИЗ ДАННЫХ АВТОБУСНОГО ПАРКА В ТАБЛИЧНОМ ПРОЦЕССОРЕ

Kulakova I.M., Lebedeva O.A., Erofeev E.V.

PROCESSING AND ANALYSIS OF BUS FLEET DATA IN A TABLE PROCESSOR

Аннотация. В статье рассматривается методика предварительной обработки статистических данных автобусного парка с использованием Microsoft Excel. Это исследование проводится на основе реальных данных транспортного предприятия. Анализ производится путем использования таких инструментов как функции для работы с данными, включая сводные таблицы, графики и встроенные формулы.

Ключевые слова: транспорт, Microsoft Excel, автобусный парк, обработка данных.

Abstract. The article discusses the methodology of preliminary processing of statistical data of the bus fleet using Microsoft Excel. This study is conducted on the basis of real data of a transport company. The analysis is performed using such tools as functions for working with data, including pivot tables, graphs and built-in formulas.

Keywords: transport, Microsoft Excel, bus fleet, data processing.

Мобильность населения является важнейшей проблемой, возникающей в процессе урбанизации. Организация городской жизни требует перемещений населения для участия в социальной и экономической деятельности. Системы общественного транспорта связаны с городской мобильностью, поскольку они представляют наиболее эффективный, устойчивый и социально доступный способ передвижения [1]. Понимание синергии между населением и общественным транспортом является ключевым фактором для повышения эффективности функционирования городского пространства. Города используют современные технологии для упрощения транспортных процессов. С умными городами связаны интеллектуальные транспортные системы (ИТС), использующие технологии для улучшения мобильности. ИТС собирают большие объемы городских данных, которые позволяют понять мобильность населения. Для этой цели анализ городских данных является цен-

ным инструментом для получения информации из необработанных источников. Понимание динамики мобильности имеет решающее значение для оптимизации транспортных систем. Мобильность описывается с помощью матриц корреспонденций, которые указывают количество пассажиров, совершающих перемещения между несколькими точками. Традиционно матрицы создаются на основе опросов или ручного подсчета пассажиров. Эти методы очень затратные для регулярного применения, и предлагают только частичное или устаревшее представление о мобильности в городе. ИТС включают технологию для определения местоположения транспортных средств и оплаты билетов в общественном транспорте. В качестве побочного продукта эти технологии генерируют ценные данные, которые можно обрабатывать для оценки матриц.

Обработку можно производить путем применения Microsoft Excel, который подходит для небольших и средних выборок и яв-

ляется отличным инструментом для анализа данных автобусного парка. Microsoft Excel, также поддерживает макросы и VBA (Visual Basic for Applications), что позволяет автоматизировать транспортные задачи и создавать требуемые функции. Программа имеет интуитивно понятный интерфейс, что позволяет ей оставаться одним из самых популярных вариантов приложений для анализа данных.

Анализ данных – это процесс сбора и обработки необработанных данных для извлечения значимой информации, которая может предоставить информацию для принятия решений. Описание процесса анализа данных представлено на рис. 1.

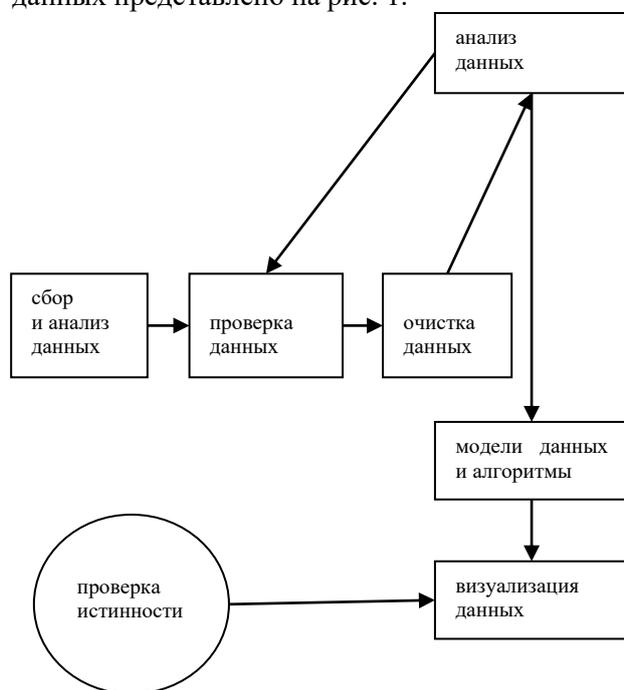


Рисунок 1 – Процесс анализа данных

Процесс анализа данных начинается и заканчивается в реальной действительности. В городских условиях это подразумевает сбор необработанных данных из определенной точки, в конечном итоге, передачу результатов в администрацию для повышения качества обслуживания населения. Процесс анализа данных традиционно состоит из нескольких этапов. Первоначально данные должны быть обработаны, включая ввод данных в таблицы, проверку наборов и их очистку для обнаружения поврежденных или неточных записей. После этого выполняется анализ данных, который направлен на описание информации полученной в процессе обработки (помимо формального мо-

делирования и проверки гипотез). Городские данные, как правило, поступают из динамичных источников (датчиков, мобильных телефонов, социальных сетей), поэтому процедура становится обязательной для поиска новых данных: посадки/высадки на остановочных пунктах; загрузки пересадочных узлов; среднего времени поездки.

Для каждого метода требуются свои атрибуты и продукты. Эти взаимодействия можно отобразить в виде ориентированного графа. При процедуре проектирования необходимо постепенно продвигаться по всей сети. Из набора доступных данных атрибуты постепенно добавляются к объекту, пока не будет достигнута цель исследования. Две основные группы методов – это методы подготовки или предварительной обработки и пошаговые методы. Последние содержат алгоритмы для реализации фактической модели транспортировки, такой как шаги распределения или назначения.

Подготовка данных представляет собой ключевой этап в рамках любого научного исследования, особенно в контексте реализации фактической модели транспортировки населения, которая включает в себя алгоритмы определения потребности в перевозках на основе реальных данных, изучение существующих потоков, а также моделирование их в различных условиях.

Этот процесс зачастую требует значительных временных затрат, что делает использование адаптированных инструментов особенно актуальным. Применение таких инструментов может существенно оптимизировать время, затрачиваемое на подготовку данных.

В этой статье представлен подход к анализу городских данных автобусного парка для изучения мобильности в табличном процессоре. В качестве примера был проанализирован и обработан реальный набор данных о местоположении автобусов и спросе по часам суток.

Важным аспектом подготовки исходных статистических данных является приведение их к сопоставимому виду. Приведение данных к сопоставимому виду обеспечивает качество и надежность данных, что, в свою очередь, влияет на точность и достоверность результатов анализа. Этот процесс помогает создать единую, структурированную и чистую базу данных, готовую к использованию для принятия обоснованных решений в об-

ласти транспортных пассажирских перевозок. Этот процесс может включать следующие операции:

- Очистка данных. Извлечение из массива заведомо недостоверных данных, выбросов и ошибок наблюдения. Здесь же может проводиться стандартизация форматов данных, устранение дублирующейся информации, а также другие методы, повышающие качество данных. Контроль единиц измерения, адекватные реакции на отсутствие данных или нетипичные значения при использовании относительных единиц – корректные базы сравнения.
- Классификация и категоризация. Присвоение данных к определённым категориям или группам, что позволяет упростить анализ (например, разделение остановочных пунктов на зоны). Создание иерархий: формирование иерархических структур для более удобного анализа данных (например, регионы, города, маршруты).
- Нормализация и масштабирование данных. Приведение данных к определённому диапазону значений, что позволяет избежать доминирования одних показателей над другими.
- Объединение данных из различных источников на основе общих ключей (например, идентификаторов маршрутов или остановок).

- Агрегация данных: Суммирование, усреднение или подсчет значений для создания сводных таблиц. Компоновка разрозненной информации в общую структуру, используемую для исследования.

Рассмотрим массив данных, полученных от транспортного предприятия.

Структура данных представляет из себя набор таблиц, каждая из которых содержит данные об определенном автобусе за определенный день и имеет следующие поля: дата, время, количество вошедших пассажиров (вход1 и вход2) и количество вышедших пассажиров (выход1 и выход2), а также данные о продолжительности стоянки.

В отдельной таблице приведены данные о времени, когда автобус должен находится на определенном остановочном пункте согласно расписанию.

После предварительной обработки данных возникает необходимость объединить данные этих двух видов таблиц, чтобы определить остановочный пункт, на котором находится автобус. Однако время, приведенное в расписании движения автобусов имеет меньшую точность, и представлено с точностью до минут, а фактические данные о движении автобусов фиксировались вплоть до секунд. Для приведения этих данных к сопоставимому виду использовался редактор Power Query (рис. 2) [3].

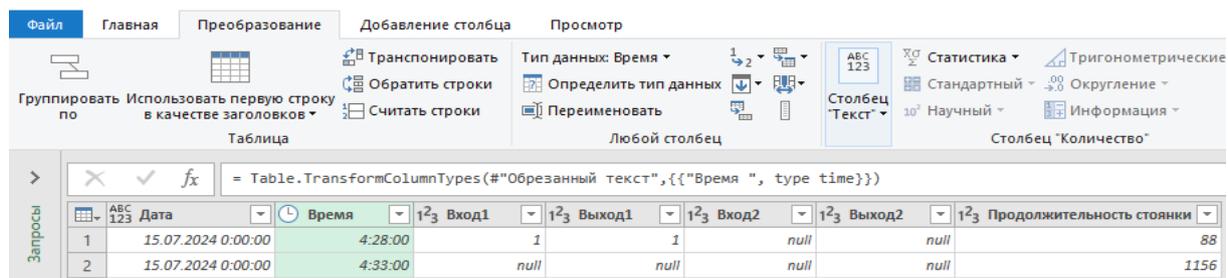


Рисунок 2 – Обработка данных в Power Query

Power Query – это мощный инструмент для извлечения, трансформации и загрузки данных, который интегрирован в Microsoft Excel и Power BI. Он позволяет пользователям подключаться к различным источникам данных, очищать и преобразовывать данные, а затем загружать их для анализа и визуализации.

Так как в реальных условиях автобус может идти с отклонением от расписания, сопоставление таблиц должно проводиться

не по полному соответствию фактического времени нахождения на остановке с временем, определенным расписанием, а с учётом ближайшего подходящего времени. Это было учтено при помощи функции Excel ВПР (VLOOKUP) с включенным интервальным просмотром. Функция позволяет выводить данные об остановке, на которой находится автобус (столбец Н рис.3) в указанное время (столбец В рис. 3) в соответствии с расписа-

нием движения автобусов (таблица Расписа-

ние диапазон \$A\$2:\$B\$55, рис. 3).

А	В	С	Д	Е	Ф	Г	Н
Дата	Время	Вход1	Выход1	Вход2	Выход2	Продолжительность стоянки	Название остановки
15.07.2024	6:26	4	1			29	ЖД вокзал
15.07.2024	6:28	1			1	6	Московский тракт (Ленина) А
15.07.2024	6:30					11	Ангарские ворота А
15.07.2024	6:32	1				5	Автостанция (ул. Кирова) А
15.07.2024	6:33	3				17	Техникум А

`=ВПР(Н4;Расписание!A2:B55;2;ИСТИНА)`

Рисунок 3 – Использование функции ВПР

Дальнейшая обработка требовала провести зонирование остановок, чтобы в дальнейшем анализировать интенсивность пассажиропотока в рамках определенных зон города. Была введена вспомогательная таблица, содержащая названия зон и остановок, данные из которой подставлялись в данные о зоне. (рис. 4).

Агрегирующие данные по остановкам с учетом зон по всем статистическим данным были получены с помощью сводных таблиц Excel (рис. 5). Сводные таблицы в Excel являются мощным инструментом для анализа и обобщения больших объемов данных. Они позволяют пользователям быстро агрегировать информацию, выявлять тенденции и проводить сравнения. С помощью сводных таблиц можно группировать данные по различным критериям, таким как даты, категории или числовые значения. Сводные таблицы поддерживают фильтрацию и сортировку, что упрощает процесс анализа.

Этот инструмент работает с большими массивами данных и позволяет прини-

мать обоснованные решения на основе полученных результатов.

А	В
Остановка	Название зоны
11-й микрорайон С	11-й микрорайон
Телекомпания Актис D	11-й микрорайон
12-й микрорайон D	12-й микрорайон
12-й микрорайон А	12-й микрорайон
179 квартал В	179 квартал
179 квартал А	179 квартал
д/у Родничок В	219-й квартал
211-й квартал В	219-й квартал
219-й квартал А	219-й квартал
29-й микрорайон В	29-й микрорайон
29-й микрорайон А	29-й микрорайон
4-й поселок	4-й поселок

Рисунок 4 – Вспомогательная таблица «Зонирование»

С помощью сводных таблиц были подведены суммарные размеры пассажиропотока по конкретным остановкам и по зонам, с возможностью фильтрации по времени, маршрутам и другим показателям.

Названия строк	Сумма по полю Вход1	Сумма по полю Вход2	Сумма по полю Вход1Вход2	Сумма по полю Выход1	Сумма по полю Выход2	Сумма по полю Выход1Выход2
11-й микрорайон	29		29		30	30
11-й микрорайон С	23		23		6	6
Телекомпания Актис D	6		6		24	24
12-й микрорайон	17		17		22	22
12-й микрорайон А	4		4		12	12
12-й микрорайон D	13		13		10	10
179 квартал	23		23	1	16	17
179 квартал А	6		6		14	14
179 квартал В	17		17	1	2	3
219-й квартал	я		я		б	б

Рисунок 5 – Сводная таблица с учётом зонирования

В результате анализа данных было получено: количество пассажиров, интервалы движения. Был проведен временной анализ в часы пик в рабочие дни. Затем пространственно-временной анализ показал, что население из удаленных районов начинает поездки раньше, чем те, кто находится ближе к объекту тяготения.

Помимо описательного использования данных ИТС для характеристики системы общественного транспорта и оценки пунктов назначения возможно применение методологии подготовки данных для построения

матриц с использованием цепочки поездок. Реализованный вариант обработки данных в исследуемом наборе, является весьма конкурентоспособным результатом.

Предлагаемый подход позволяет изучать мобильность на более мелком уровне. Реализованное решение является недорогим, если инфраструктура ИТС уже развернута, и позволяет предоставлять актуальную информацию о мобильности, поддерживает работу на разных уровнях агрегации и предлагает несколько инструментов для фильтрации данных.

СПИСОК ЛИТЕРАТУРЫ

1. **Лебедева, О.А.** Сравнительный анализ методов решения транспортных задач при оптимальном планировании перевозочного процесса /О.А. Лебедева., В.Е. Гозбенко, А.А. Пыхалов, Ю.Ф. Мухопад // Современные технологии. Системный анализ. Моделирование. 2020. № 3 (67). С. 134-139.

2. **Уэйн, Винстон** Бизнес-моделирование и анализ данных. Решение актуальных задач с помощью Microsoft Excel. 6-е издание. – СПб.: Питер, 2021. – 944 с.

3. **Деклер, Г.** Power Query и язык M. Подробное руководство. Гроот Р.; Корте М. – ДМК Пресс, 2024. – 746 с.