МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Рабочая программа дисциплины (модуля)
ХРАНИЛИЩА ДАННЫХ

Направление и направленность (профиль)
11.03.02 Инфокоммуникационные технологии и системы связи. Интернет-вещей и оптические системы и сети

Год набора на ОПОП
2024

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Хранилища данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 11.03.02 Инфокоммуникационные технологии и системы связи (утв. приказом Минобрнауки России от 19.09.2017г. №930) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Кригер А.Б.

Утверждена на заседании кафедры информационных технологий и систем от 28.05.2026 , протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	0000000000F796F2
Владелец	Кийкова Е.В.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Цели освоения учебной дисциплины:

- получение базовых знаний о системах хранения данных**,** особенностях Хранилищ данных и их назначении;
- формирование умений и навыков проектирования Хранилищ данных для систем бизнес-анализа;
- знакомство с технологиями интеллектуального анализа.

Задачи освоения дисциплины:

- изучение принципов проектирования и разработки хранилищ данных;
- получение навыков настройки хранилищ данных;
- проектирование и разработка процесса наполнения Хранилища данных, реализации запросов к Хранилищам данных;

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисципл		
			Код результата	Формулировка результ	
11.03.02 «Инфокоммуникационные технологии и системы связи» (Б-ИК)	ПКВ-4 : Способен осуществлять сбор и анализ исходных данных для проектирования и проектировать устройства связи, интеллектуальные инфокоммуникационные сети и их элементы	ПКВ-4.1к : Осуществляет сбор технических данных для проектирования телекоммуникационных устройств	РД1	Знание	теоретических основ многомерной модели данны
			РД2	Умение	использовать системы хран данных, соответствую задачам профессионал деятельности
			РД3	Навык	разработки логических моделей хранилищ дан

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	**Формирование ценностей**	**Целевые ориентиры**
Формирование гражданской позиции и патриотизма		
Развитие патриотизма и гражданской ответственности	Созидательный труд	Ответственность Активная жизненная позиция
Формирование духовно-нравственных ценностей		

Формирование ответственного отношения к труду	Взаимопомощь и взаимоуважение	Дисциплинированность
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Созидательный труд	Системное мышление Гибкость мышления
Формирование коммуникативных навыков и культуры общения		
Развитие умения эффективно общаться и сотрудничать	Коллективизм	Креативное мышление Инициативность

2 Место дисциплины (модуля) в структуре ОПОП

Освоение дисциплины формирует у обучающихся компетенции, необходимые для подготовки бакалавра в соответствии с требованиями ФГОС ВО.

Дисциплина относится к элективным дисциплинам учебного плана.

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость	Объем контактной работы (час)						СРС	Форма аттестации
				(З.Е.)	Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА	КСР		
11.03.02 Инфокоммуникационные технологии и системы связи	ОФО	Б1.ДВ.Б	6	3	55	18	36	0	1	0	53	ДЗ

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Архитектуры данных: история развития.	РД1	2	0	0	2	практическое задание № 1

2	Архитектуры данных: базы данных и модели данных. Многомерные данные. Концепция хранилищ данных.	РД1, РД2	8	10	0	20	практическое задание № 1, 2-5
3	Архитектуры хранилищ данных. Реляционные хранилища данных. Реализация реляционных хранилищ данных.	РД2	4	8	0	8	практическое задание № 6-7
4	Проектирование хранилищ данных для выбранной предметной области	РД3	4	18	0	23	практическое задание № 8-11. Индивидуальное задание «Проектирование и разработка ХД».
Итого по таблице			**18**	**36**	**0**	**53**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Архитектуры данных: история развития.

Содержание темы: Эволюция задач сбора и обработки информации. Понятие архитектуры данных. Развитие систем хранения и обработки данных. Системы оперативной обработки информации – OLTP. Системы консолидации и аналитической обработки информации – ELT.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция.

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 2 Архитектуры данных: базы данных и модели данных. Многомерные данные. Концепция хранилищ данных.

Содержание темы: Иерархическая модель данных, условия целостности иерархической модели данных. Сетевая модель данных, условия целостности сетевой модели данных. Реляционная модель данных, реляционные базы данных. Хранилища данных – системы хранения данных, ориентированные на аналитическую обработку данных. OLAP-технология как ключевой компонент хранилищ данных. Задачи OLAP-систем: представление данных, визуализация данных, методы обработки данных. Концепция многомерного представления данных – гиперкубы. Базовые понятия: измерения и факты. Формализация многомерного представления данных: метки, иерархии, ячейки, меры. Построение информационных систем на основе архитектур хранилищ данных. Операции над многомерными данными. Методы обработки агрегированных данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические занятия, метод активного обучения – «мастер-класс», компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 3 Архитектуры хранилищ данных. Реляционные хранилища данных. Реализация реляционных хранилищ данных.

Содержание темы: Современное представление. Классификация архитектур данных. Многомерные реляционные и гибридные хранилища данных. Различие концепций и особенности построения. Применение реляционной модели для создания хранилищ данных. Архитектуры реляционных хранилищ данных: «звезда», снежинка». Особенности реализации реляционных хранилищ данных. Анализ и преобразование исходных данных.

Выбор архитектуры реляционных хранилищ данных. Метаданные. Этапы реализации проекта. Реализация аналитической обработки данных, загруженных в хранилища данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические занятия, компьютерное моделирование.

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 4 Проектирование хранилищ данных для выбранной предметной области.

Содержание темы: Виртуальные хранилища данных. Использование хранилищ данных. Различные архитектурные решения хранилищ данных, реализация процедур ETL. Анализ данных предметной области для загрузки в хранилище данных. Разработка модели хранилища данных. Реализация реляционного хранилища данных. Загрузка данных в хранилище данных. Проверка работоспособности. Выбор методов аналитической обработки данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практические занятия, компьютерное моделирование, метод активного обучения – «конференция».

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

Современные корпоративные информационные системы содержат приложения, предназначенные для комплексного анализа данных, поддержки принятия решения, извлечения знаний.

Принятия эффективного управленческого решения невозможно, без консолидированных отчетов и глубинного анализа результатов. Для этого необходимо создание хранилищ данных (Data warehouses), специальных систем хранения, обеспечивающих загрузку предварительно извлеченных и обработанных данных и оперативный доступ к ним.

Основные требования к хранилищам данных:

- поддержка высокой скорости доступа к данным;
- поддержка внутренней непротиворечивости данных;
- возможность манипулирования данными;
- полнота и достоверность хранимых данных;
- поддержка процессов обновления данных.

В отличие от оперативных баз данных, на основе которых строятся учетные информационные системы, хранилища данных предназначены исключительно для аналитической обработки данных. Данные загружаются в хранилище из оперативных баз данных.

Перечень и тематика самостоятельных работ студентов по дисциплине

Тема 1.

1) Работа с платформой Loginom Academic 6.2.1. Освоение загрузки данных и методов визуализации.

2) Методы предварительной обработки данных, реализованные в Loginom Academic 6.2.1. Рассмотреть назначение с методов, составить таблицу, отражающую технологию предварительной обработки и соответствующий математический метод.

3) Аналитическая обработка данных средствами Loginom Academic 6.2.1.

Тема 2.

1) Изучение учебной РХД (поставляется разработчиком платформы).

Тема 3

1) Разработка ER-модели для заданной предметной области.

Тема 4.

1) Анализ предметной области (для разработки витрины данных).

2) Подготовка данных из открытых источников – извлечение исходных данных для дальнейшей загрузки в реляционное хранилище данных.

3) Разработка логической модели реляционного хранилища данных для выбранной предметной области.

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Агальцов, В. П. Базы данных : учебник : в 2 книгах. Книга 1. Локальные базы данных / В.П. Агальцов. — Москва : ФОРУМ : ИНФРА-М, 2025. — 352 с. : ил. — (Высшее образование). - ISBN 978-5-8199-0377-3. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2162084 (дата обращения: 31.05.2026)

2. Дадян, Э. Г. Данные: хранение и обработка : учебник / Э.Г. Дадян. — Москва : ИНФРА-М, 2026. — 205 с. — (Высшее образование). — DOI 10.12737/textbook_5cf8c7f2b8cdb8.06963680. - ISBN 978-5-16-021135-0. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2214875 (дата обращения: 31.05.2026)

7.2 Дополнительная литература

1. Бененсон, М. З. Информационные системы, базы и хранилища данных : методические указания / М. З. Бененсон, С. А. Сорокин. — Москва : РТУ МИРЭА, 2024. — 37 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/405209 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

2. Березовская, Е. А. Работа с аналитической платформой Loginom : учебное пособие / Е. А. Березовская, С. В. Крюков ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2024. - 100 с. - ISBN 978-5-9275-4731-9. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2191408 (дата обращения: 31.05.2026)

3. Смирнов, М. В. Проектирование хранилищ данных : учебно-методическое пособие / М. В. Смирнов. — Москва : РТУ МИРЭА, 2024. — 91 с. — ISBN 978-5-7339-2258-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/432674 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Электронно-библиотечная система "ZNANIUM.COM"
2. Электронно-библиотечная система "ЛАНЬ"
3. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
4. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/
5. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Ист.бесп.эл.питания Smart-UPS 3000VA
- Источник б/переб. пит. №2 APC SmartUPS 3000
- Монитор облачный 23" LG23CAV42K/мышь Genius Optical Wheel проводная/клавиатура Genius KB110 проводная
- Мультимедийный комплект №2 в составе:проектор Casio XJ-M146,экран 180*180,крепление потолочное
- Облачный монитор 23" LG CAV42K
- Система аудиовизуального представления информации
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Microsoft Office Professional Plus 2013 Russian
- Microsoft OfficeProffessionalPlus 2019 Russian
- Microsoft SQL Server Standard 2008R2 Russian

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

ХРАНИЛИЩА ДАННЫХ

Направление и направленность (профиль)
11.03.02 Инфокоммуникационные технологии и системы связи. Интернет-вещей и оптические системы и сети

Год набора на ОПОП
2024

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
11.03.02 «Инфокоммуникационные технологии и системы связи» (Б-ИК)	ПКВ-4 : Способен осуществлять сбор и анализ исходных данных для проектирования и проектировать устройства связи, интеллектуальные инфокоммуникационные сети и их элементы	ПКВ-4.1к : Осуществляет сбор технических данных для проектирования телекоммуникационных устройств

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

Компетенция ПКВ-4 «Способен осуществлять сбор и анализ исходных данных для проектирования и проектировать устройства связи, интеллектуальные инфокоммуникационные сети и их элементы»

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ПКВ-4.1к : Осуществляет сбор технических данных для проектирования телекоммуникационных устройств	РД 1	Знание	теоретических основ многомерной модели данных	сформировавшееся знание теоретических основ многомерной модели данных знание
	РД 2	Умение	использовать системы хранения данных, соответствующие задачам профессиональной деятельности	сформировавшееся умение использовать системы хранения данных, соответствующие задачам профессиональной деятельности
	РД 3	Навык	разработки логических моделей хранилищ данных	сформировавшееся владение навыками разработки хранилищ данных

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения	Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС

			Текущий контроль	Промежуточная ат тестация
Очная форма обучения				
РД1	Знание : теоретических основ многомерной мод ели данных	1.1. Архитектуры данны х: история развития.	Тест	Проект
		1.2. Архитектуры данны х: базы данных и модел и данных. Многомерные данные. Концепция хра нилищ данных.	Тест	Проект
РД2	Умение : использовать с истемы хранения данны х, соответствующие зад ачам профессиональной деятельности	1.2. Архитектуры данны х: базы данных и модел и данных. Многомерные данные. Концепция хра нилищ данных.	Тест	Проект
		1.3. Архитектуры храни лищ данных. Реляционн ые хранилища данных. Реализация реляционны х хранилищ данных.	Тест	Проект
РД3	Навык : разработки логи ческих моделей хранили щ данных	1.4. Проектирование хра нилищ данных для выбр анной предметной облас ти	Тест	Проект

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Оценочное средст во		
	Тест (по результат ам практических з аданий 1-7)	Индивидуальное задани е – проект (практически е задания 8 -11)	Итого
Лекции	20		20
Практические занятия		30	30
Самостоятельная работа	20		20
Промежуточная аттестация		30	30
Итого	40	60	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма балло в по дисципли не	Оценка по промеж уточной аттестаци и	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обна руживает всестороннее, систематическое и глубокое знание учебного материала , усвоил основную литературу и знаком с дополнительной литературой, рекоме ндованной программой, умеет свободно выполнять практические задания, пред усмотренные программой, свободно оперирует приобретенными знаниями, уме ниями, применяет их в ситуациях повышенной сложности.
от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: осно вные знания, умения освоены, но допускаются незначительные ошибки, неточн

		ости, затруднения при аналитических операциях, переносе знаний и умений на н овые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворитель но»	Студент демонстрирует сформированность дисциплинарных компетенций: в хо де контрольных мероприятий допускаются значительные ошибки, проявляется о тсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным к омпетенциям, студент испытывает значительные затруднения при оперировани и знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворите льно»	У студента не сформированы дисциплинарные компетенции, проявляется недос таточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворите льно»	Дисциплинарные компетенции не сформированы. Проявляется полное или прак тически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Темы групповых и/или индивидуальных проектов

Проектирование и разработка виртуального ХД средствами loginom (требования к проектному заданию)

1. Провести анализ массива данных (с позиции ведения бизнеса и экономики): выделить объекты и их характеристики, выделить показатели результатов деятельности.
2. Определить метаданные гиперкуба: измерения – типы данных и форматы представления, факты – типы данных, форматы представления, методы агрегации, число агрегированных показателей.
3. Сформировать бизнес-метаданные: атрибуты объектов, показатели предметной области, возможные методы обработки, экспертные заключения.
4. Загрузить данные в loginom, формировать единый массив данных (dataframe), использовать реляционные связи между отдельными таблицами. Сформирование OLAP-кубы, и срезы OLAP-кубов (элемент «кросс-таблица»).
5. Провести корреляционный анализ для показателей-фактов, средствами loginom построить модели регрессии, нейросетевые модели, объясняющие изменение переменной-признака.

Примерные темы проектных заданий - предметные области для проектирования хранилищ данных:

1. Исследование рынка (глобального): мировые рынки продовольствия (зерновые, кофе, молочная продукция), мировые и региональные рынки металлов, мировые и локальные рынки смартфонов и планшетов, компьютеров и т.д.;

2. Рынок подержанных автомобилей;

3. Рынок водного транспорта (катера, аквабайки, скутеры и т.д.);

4. Рынок жилья: новостройки, реализация от застройщика не сданного в эксплуатацию фонда;

5. Цены, ценообразование, динамика ценовых уровней: продажи продовольственных товаров, продажи бытовой техники, продажи стройматериалов, цены на услуги, цены на компьютерную технику, цены мобильные телефон, смартфоны;

6. Результаты маркетинговых исследований (по регионам или отраслям);

7. Эффективность систем электронной коммерции: посещаемость, продажи, обратная связь, отзывы;

8. Управление складом;

9. Региональные бюджеты, бюджеты районов и округов.

Краткие методические указания

Индивидуальное задание выполняется как на практических занятиях (раздел «Проектирование ХД»), так и в рамках СРС. После выполнения индивидуального задания результаты проекта предоставляются на проверку преподавателю. В состав материалов для проверки входят: исходные данные в форматах табличного процессора или текстовом, ХД реализуется средствами Loginom, пояснительная записка.

Шкала оценки

№	Баллы	Описание
5	36-40	Студент выполнил индивидуальное задание в соответствии с требованиями, демонстрирует способность выбрать тип архитектуры хранилища данных, построить модель данных, выбрать метод консолидации данных. Реализовал работоспособное локальное хранилище данных (витрину данных). Обосновал выбор методов обработки данных.
4	26-35	Студент демонстрирует умения на среднем уровне. Демонстрирует способность выбрать тип архитектуры хранилища данных, построить модель данных. Реализовал работоспособное локальное хранилище данных (витрину данных). Испытывает затруднения в выборе и обосновании выбора методов обработки данных.
3	16-25	Студент демонстрирует умения и навыки на базовом уровне. Способен выбрать тип архитектуры хранилища данных, построить модель данных. Реализовал работоспособное локальное хранилище данных (витрину данных), однако не сумел построить запросы, реализующие сечения многомерного куба. Испытывает затруднения выборе и обосновании выбора методов обработки данных.
2	8-15	Студент демонстрирует умения и навыки на уровне ниже базового. Испытывает затруднения в выборе архитектуры хранилища данных, в построении модели данных.
1	0-7	Студент проявляет полное или практически полное отсутствие умений и навыков, но присутствует на занятиях и пытается выполнить задание.

5.2 Контрольный тест

Предусмотрены два вида тестов:

1. тест, направленный на проверку теоретических знаний

2. тест, направленный на проверку умений и навыков создавать систему хранения и выполнять аналитическую обработку с помощью аналитической платформы Loginom.

Последний тест предполагает создание локального виртуального ХД, и только потом - ответы на вопросы.

Т.к. тест типа 2 очень громоздок в представлении, ОН ЗАГРУЖЕН НЕПОСРЕДСТВЕННО ВМЕСТЕ С КЛЮЧАМИ ФОС.

Краткие методические указания

1. Для выполнения контрольных тестов необходимо освоить Знания в области разработки систем хранения данных

2. Изучить теоретические основы разработки ХД

3. Выполнить Практические задания:

1. Изучить инструкции по работе с Loginom.
2. Освоить реализацию локальных виртуальных ХД средствами Loginom.
3. **Выполнить практические задания по темам курса. Использовать инструкции**

Шкала оценки

№	Баллы	Описание
5	16-20	Студент демонстрирует умения и знания на высоком уровне. Реализовал работоспособное локальное хранилище данных (витрину данных). Ответил на вопросы теста
4	11-15	Студент демонстрирует умения на среднем уровне. Реализовал работоспособное локальное хранилище данных (витрину данных). Испытывает затруднения в выборе и обосновании выбора методов обработки данных. Ответил на вопросы теста.
3	6-10	Студент демонстрирует умения и навыки на базовом уровне. Реализовал работоспособное локальное хранилище (витрину данных), однако не сумел построить запросы, реализующие сечения многомерного куба.
2	0-5	Студент демонстрирует умения и навыки на уровне ниже базового. Не способен ответить на вопросы теста.

Банк тестовых заданий по дисциплине «Хранилища данных»

Блок 1

Тест базовых вопросов теме: "Хранение и предварительная обработка данных"

Формат: 10 вопросов (закрытые, открытые)

1. Что такое хранилище данных (Data Warehouse)?
a) База данных для оперативной обработки транзакций
b) Централизованное хранилище структурированных данных для анализа
c) Файловое хранилище для резервных копий
d) Облачное хранилище неструктурированных данных

2. Какие из перечисленных процессов относятся к предварительной обработке данных? (Выберите несколько вариантов)
a) Нормализация данных
b) Шифрование данных
c) Заполнение пропущенных значений
d) Удаление выбросов

3. Какой подход в хранилищах данных предполагает хранение информации в виде «сырых» и агрегированных данных?
a) Схема «звезда»
b) Data Vault
c) Слоистая архитектура (RAW → CLEAN → BUSINESS)
d) Денормализованная таблица

4. Для чего используется ETL-процесс в хранилищах данных?
a) Для визуализации данных
b) Для извлечения, преобразования и загрузки данных
c) Для защиты данных от взлома
d) Для удаления избыточных данных

5. Какая операция предварительной обработки помогает устранить дублирование записей?
a) Дискретизация
b) Дедупликация
c) Нормализация
d) Кластеризация

6. Какой метод предобработки данных используется для приведения признаков к единому масштабу?
a) Кодирование категориальных переменных
b) Стандартизация (Z-score)
c) Удаление шумов

d) Импутация данных

7. В чём основное отличие OLAP от OLTP?
a) OLAP используется для оперативных транзакций, OLTP — для аналитики
b) OLAP оптимизирован для аналитических запросов, OLTP — для обработки транзакций
c) OLAP работает только с неструктурированными данными
d) OLTP не поддерживает SQL-запросы

8. Какие проблемы решает предварительная обработка данных? (Открытый вопрос)

9. Перечислите 3 слоя архитектуры хранилища данных, хаарктериные для современных архитектур. (Краткий ответ)

10. Какой инструмент или метод вы бы использовали для обработки пропущенных значений в данных? Обоснуйте выбор. (Развёрнутый ответ)

Ответы:
1. b; 2. a c d; 3. d; 4. b; 5. b;
6. b; 7. b;
8. Предобработка данных необходима их подготовка к анализу, в соответствии с выбранными методами и моделями анализа, определяемыми спецификой решаемой задачи
9. Слой загрузки , слой хранения, слой витрин данных
10. Здесь ваш творчески осмысленный ответ:
Критерии оценки:
1-7 вопросы (закрытые) — 1 балл каждый.
8-9 вопросы (краткие) — 2 балла каждый.
10 вопрос (развёрнутый) — 3 балла.
Максимум: 14 баллов.

Блок 2
Часть I: Вопросы с выбором одного правильного ответа
1. Что такое хранилище данных согласно тексту?
a) Система для автоматизации бизнес-процессов
b) Предметно-ориентированный, интегрированный, зависимый от времени набор данных
c) Инструмент для анализа больших объемов данных
d) Репозиторий слабо структурированных данных

2. Какой из следующих факторов НЕ относится к предпосылкам создания хранилищ данных?
a) Ужесточение конкуренции
b) Развитие CRM-систем
c) Централизованное управление данными

d) Разрозненность данных

3. Что из следующего верно относительно процедуры загрузки данных в хранилище?
a) Процесс загрузки включает копирование данных
b) Процесс загрузки включает очистку, согласование и контроль качества данных
c) Процесс загрузки требует минимальных ресурсов
d) Процесс загрузки возможен без предварительного определения целей

4. Как называются инструменты, позволяющие пользователям получать быстрый доступ к данным для формирования запросов?
a) OLTP
b) OLAP
c) Ad-hoc reporting
d) Data mining

5. Что из нижеперечисленного является основным отличием хранилища данных от оперативной базы данных?
a) Хранилище данных ориентировано на анализ, а не на автоматизацию
b) Хранилище данных содержит исторические данные
c) Хранилище данных использует многомерную модель данных
d) Хранилище данных централизует данные из разных источников

Часть II: Вопросы с выбором нескольких правильных ответов

6. Какие из следующих утверждений верны относительно хранилищ данных?
a) Они используются для анализа данных, а не для автоматизации бизнес-процессов
b) Могут быть реализованы как на реляционных, так и на многомерных СУБД
c) Всегда строятся с использованием OLAP инструментов
d) Требуют тесного сотрудничества между ИТ-специалистами и бизнес-подразделениями

7. Какие из следующих проблем могут возникнуть при неправильном проектировании хранилища данных?
a) Перегруженность ненужными данными
b) Неправильное использование ресурсов
c) Неполная востребованность данных
d) Сложности с интеграцией данных из различных источников

Часть III: Вопрос на различие между Базой данных и Хранилищем данных

8. Укажите основные отличия Базы данных от Хранилища данных:
a) База данных ориентирована на автоматизацию бизнес-процессов, тогда как хранилище данных – на анализ данных
b) База данных содержит текущие данные, тогда как хранилище данных – исторические
c) База данных централизована, тогда как хранилище данных распределено

d) База данных использует транзакционную модель обработки данных, тогда как хранилище данных – аналитическую

Ответы на вопросы блока 2
1. b
2. c
3. b
4. c
5. a
6. a, b, d
7. a, c, d
8. a, b, d

Блок 3
Тест по дисциплине «Хранилища данных»
Формат: 40 вопросов (выбор одного ответа), разделённых на 4 тематических блока.

Раздел 1. Основы хранилищ данных
1. Что такое хранилище данных (Data Warehouse)?
a) Система для оперативной обработки транзакций (OLTP).
b) Централизованное хранилище для аналитики и отчётности.
c) База данных для хранения неструктурированных документов.
d) Сервис для резервного копирования файлов.

2. Какой подход к проектированию хранилища использует слои RAW → CLEAN → BUSINESS?
a) Data Vault.
b) Снежинка (Snowflake).
c) Слоистая архитектура.
d) Денормализованная схема.

3. Какой тип данных НЕ характерен для хранилищ?
a) Метаданные.
b) Агрегированные данные.
c) Временные ряды.
d) Текстовые заметки пользователей.

4. Что означает принцип «Subject-Oriented» в хранилищах данных?
a) Данные организованы по тематическим областям (продажи, клиенты).
b) Данные хранятся в хронологическом порядке.
c) Все данные нормализованы.
d) Хранилище доступно только определённым пользователям.

5. Какой процесс НЕ входит в задачи хранилища данных?

a) Поддержка OLAP-запросов.
b) Репликация данных в реальном времени.
c) Хранение исторических данных.
d) Интеграция данных из разных источников.

6. Что такое «озеро данных» (Data Lake) в сравнении с хранилищем?
a) Хранилище только для структурированных данных.
b) Система для хранения сырых данных любого формата.
c) Инструмент для визуализации данных.
d) База данных с поддержкой ACID-транзакций.

7. Какой стандарт метаданных используется в хранилищах?
a) JSON.
b) XML.
c) ISO 11179.
d) CSV.

8. Что такое «витрина данных» (Data Mart)?
a) Подмножество хранилища для конкретного отдела.
b) Инструмент для очистки данных.
c) Система шифрования данных.
d) База данных для IoT-устройств.

9. Какой язык запросов чаще всего используется для аналитики в хранилищах?
a) SQL.
b) Python.
c) NoQL.
d) R.

10. Что означает «Non-Volatile» в контексте хранилищ?
a) Данные постоянно обновляются.
b) Данные защищены от удаления.
c) Данные доступны только для чтения.
d) Данные не изменяются после загрузки.

Раздел 2. ETL и интеграция данных
11. Что означает аббревиатура ETL?
a) Extract, Transform, Load.
b) Export, Transfer, Loop.
c) Encrypt, Tag, Log.
d) Evaluate, Test, Launch.

12. Какой этап ETL отвечает за фильтрацию дубликатов?
a) Extract.
b) Transform.
c) Load.

d) Все перечисленные.

13. Что такое CDC (Change Data Capture)?
a) Метод шифрования данных.
b) Фиксация изменений в источниках для ETL.
c) Инструмент визуализации.
d) Протокол передачи данных.

14. Какой инструмент НЕ является ETL-решением?
a) Apache NiFi.
b) Talend.
c) Tableau.
d) Informatica PowerCenter.

15. Зачем нужен «Staging Area» в ETL?
a) Для хранения сырых данных перед обработкой.
b) Для публикации отчётов.
c) Для резервного копирования.
d) Для тестирования SQL-запросов.

Ответы на вопросы блока 3
1. b; 2. c; 3. d; 4. a; 5. b; 6. b; 7. c; 8. a; 9. a;
10. d; 11. a; 12. b; 13. b; 14. c; 15. a

Блок 4 – вопросы повышенного уровня сложности
Раздел 3. Предварительная обработка данных
Вопросы с выбором одного или нескольких ответов

1. Какие из перечисленных этапов относятся к предварительной обработке данных? (Выберите все верные варианты)
1. Очистка данных от шумов
2. Заполнение пропущенных значений
3. Шифрование данных
4. Нормализация данных
5. Развёртывание модели машинного обучения

Ответ: 1, 2, 4

2. Какой метод используется для обработки пропущенных значений в данных? (Выберите все подходящие)
1. Удаление строк с пропусками
2. Заполнение средним/медианным значением
3. Использование алгоритмов предсказания (например, KNN)
4. Шифрование пропущенных полей

Ответ: 1, 2, 3

3. Что такое «выбросы» (outliers) в данных?
1. Значения, значительно отклоняющиеся от общего распределения
2. Ошибки в формате данных
3. Пропущенные значения
4. Дублирующиеся записи

Ответ: 1

4. Какие методы помогают обнаружить выбросы? (Выберите все верные)
1. Метод межквартильного размаха (IQR)
2. Z-score стандартизация
3. One-Hot Encoding
4. Визуализация (например, boxplot)

Ответ: 1, 2, 4

5. Для чего применяется нормализация данных?
1. Приведение признаков к единому масштабу
2. Удаление дубликатов
3. Шифрование конфиденциальных данных
4. Сжатие данных

Ответ: 1

6. Какие методы нормализации существуют? (Выберите все подходящие)
1. Min-Max Scaling
2. Z-score стандартизация
3. PCA (Principal Component Analysis)
4. Decimal Scaling

Ответ: 1, 2, 4

7. Что такое «бининг» (binning) в предобработке данных?
1. Группировка числовых значений в интервалы
2. Удаление столбцов с высоким уровнем пропусков
3. Преобразование категориальных данных в числовые
4. Автоматическое исправление опечаток

Ответ: 1

8. Какие подходы используются для обработки категориальных данных? (Выберите все верные)
1. One-Hot Encoding
2. Label Encoding

3. Z-score преобразование
4. Target Encoding

Ответ: 1, 2, 4

9. Какой метод НЕ применяется для очистки текстовых данных?
1. Удаление стоп-слов
2. Лемматизация
3. Нормализация по методу Min-Max
4. Исправление опечаток

Ответ: 3

10. Что такое «дедупликация»?
1. Удаление дублирующихся записей
2. Объединение нескольких столбцов в один
3. Замена пропусков нулями
4. Автоматическое создание метаданных

Ответ: 1

11. Какой инструмент используется для предобработки данных в Loginom? (Выберите все подходящие)
1. Визуальные компоненты для ETL
2. Скрипты на Python/R
3. Нейросетевые модели
4. Встроенные алгоритмы очистки

Ответ: 1, 2, 4

12. Зачем нужна обработка несбалансированных данных?
1. Чтобы избежать смещения (bias) в моделях машинного обучения
2. Для уменьшения размера датасета
3. Для шифрования данных
4. Для ускорения SQL-запросов

Ответ: 1

13. Какие методы помогают бороться с несбалансированностью? (Выберите все верные)
1. Oversampling (например, SMOTE)
2. Undersampling
3. Удаление всех аномалий
4. Использование взвешенных классов

Ответ: 1, 2, 4

Критерии оценки для раздела:
Вопросы с одним ответом: 1 балл.
Вопросы с множественным выбором: 2 балла (если все верные варианты отмечены).

Тесты для проверки умений и навыков работы с аналитической платформой Loginom

Тест 1. Многомерные данные, OLAP – кубы, Loginom

Используется много мерный датафрейм Ships. Для получения ответов необходимо создать проект в Loginom, выстроить верный сценарий, создать OLAP-кубы. Правильные ответы выделены цветом

1. Укажите все факты которые могут быть определены для датафрейма Ships (с позиции изучения характеристик \ цены)
   a) Название
   b) Цена
   c) Ширина
   d) Осадка
   e) Скорость
   f) Запас топлива
   g) Вес
   h) Пассажировместимость
   i) страна

2. Создано сечение гиперкуба – пассажировместимость \ страна. Факты: вес, цена. Метод агрегации среднее. Укажите правильные элементы сечения куба:

   a) (Норвегия/3) вес=1988, цена=867000;(Россия/8) вес=2590,цена=3194694
   b) (Норвегия/3) вес=867000, цена=1988;(Россия/8) вес=2590,цена=3194694
   c) (Норвегия/3) вес=1176, цена=20000;(Россия/8) вес=2590,цена=4757834
   d) (Норвегия/3) вес=2072, цена=2012860;(Россия/8) вес=2122,цена=2159896

3. Создано сечение гиперкуба – класс мореходности\ год выпуска. Факт: вес. Методы агрегации минимальное \ максимальное. Укажите ВСЕ правильные элементы сечения куба:

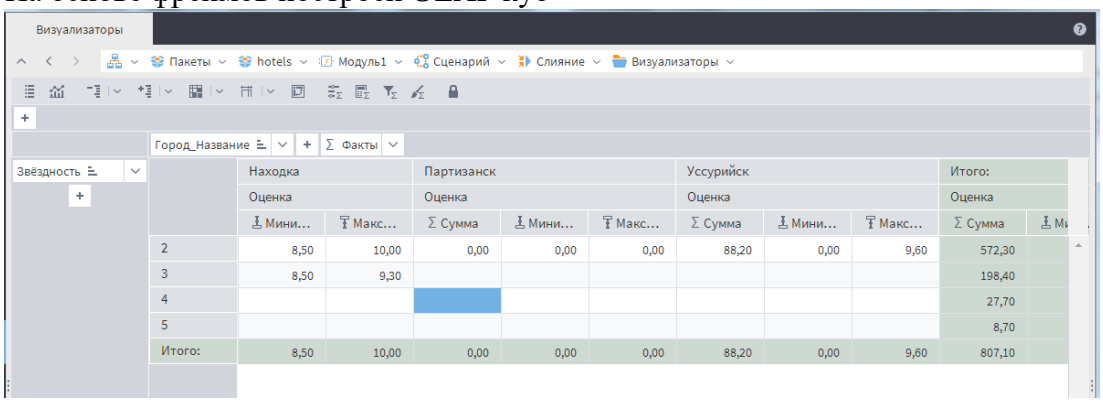
   a) (D / 2013) весmin= 2128, весmax=2128
   b) (D / 2013) весmin= 1064, весmax=2968
   c) (A / 2015) весmin= 2128, весmax=3584
   d) (B / 2015) весmin= 1064, весmax=2380
   e) (C / 2015) весmin= 1064, весmax=2800

Дан дата фрейм, представленный массивами (5 файлов): «Отели», «Факты», «Города» и т.п.
Выполните необходимы преобразования и ответьте на вопросы. ВЫБЕРИТЕ ВСЕ ПРАВИЛЬНЫЕ ОТВЕТЫ! – ОТВЕТ МОЖЕТ БЫТЬ ОДНИ , МОЖЕТ БЫТЬ БОЛЬШЕ ОДНОГО!

1. Укажите в каких файлах содержатся «измерения» (ответ поясните)
   a) «Отели», «Факты»
   b) «Факты», «Города»
   c) «Отели», «Города»

d) «Отели», «Факты», «Города»
e) «Города»
f) «Отели»

На основе фреймов построен OLAP куб

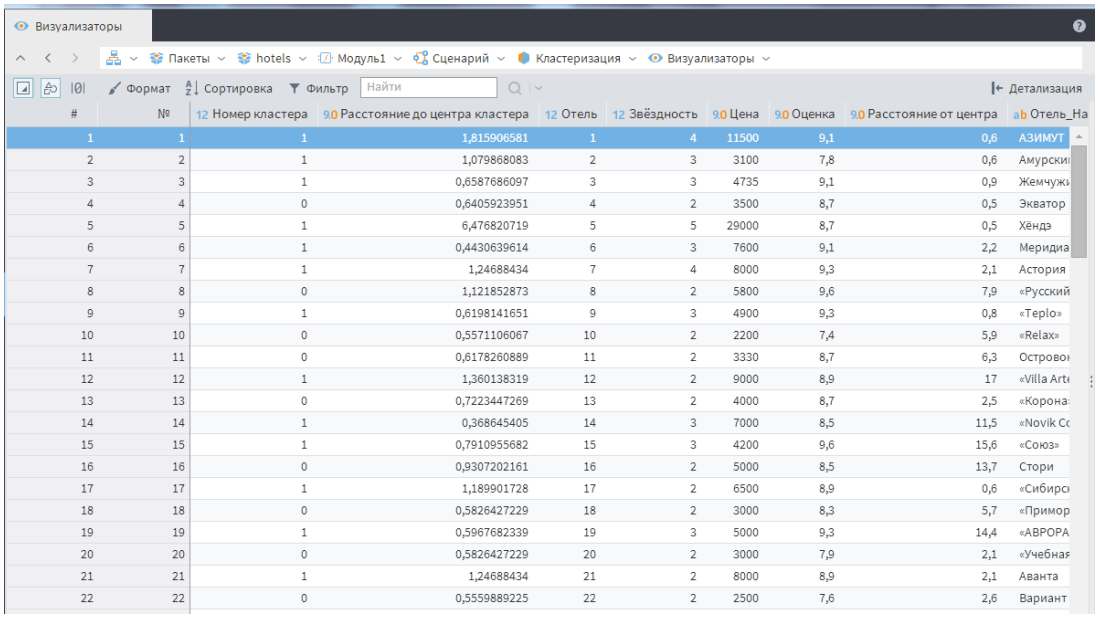


2. Укажите, какие таблицы (файлы) использованы для построения куба
a) «Отели», «Факты»
b) «Факты», «Города»
c) «Отели», «Города»
d) «Отели», «Факты», «Города»
e) «Города»
f) «Отели»

3. Укажите какой вид трансформации использовался, чтобы получить такой куб:
a) Сортировка
b) Группировка
c) Слияние
d) Кросс-таблица

4. На скринщоте результат разбиения на кластеры (3 кластера, EM кластеризация). Какие входные параметры использованы для операции? Выберите все ответы!
a) Отель
b) Звездность
c) Расстояние до центра
d) Цена
e) Оценка
f) Отель _название

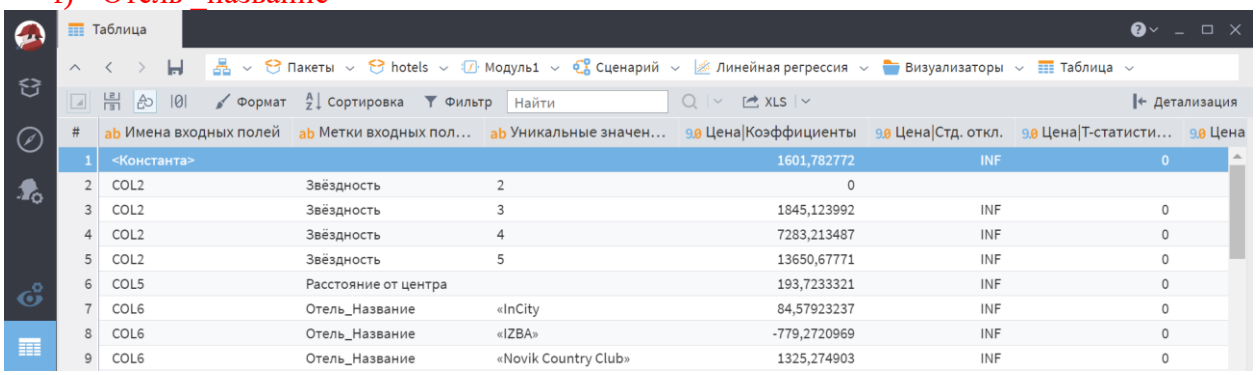
Ответить правильно сможете в двух случаях: или хорошо понимаете алгоритм, или построите куб.



#	№	Номер кластера	Расстояние до центра кластера	Отель	Звёздность	Цена	Оценка	Расстояние от центра	Отель_На
1	1	1	1,815906581	1	4	11500	9,1	0,6	АЗИМУТ
2	2	1	1,079868083	2	3	3100	7,8	0,6	Амурски
3	3	1	0,6587686097	3	3	4735	9,1	0,9	Жемчужи
4	4	0	0,6405923951	4	2	3500	8,7	0,5	Экватор
5	5	1	6,476820719	5	5	29000	8,7	0,5	Хёндэ
6	6	1	0,4430639614	6	3	7600	9,1	2,2	Меридиа
7	7	1	1,24688434	7	4	8000	9,3	2,1	Астория
8	8	0	1,121852873	8	2	5800	9,6	7,9	«Русский
9	9	1	0,6198141651	9	3	4900	9,3	0,8	«Teplo»
10	10	0	0,5571106067	10	2	2200	7,4	5,9	«Relax»
11	11	0	0,6178260889	11	2	3330	8,7	6,3	Островок
12	12	1	1,360138319	12	2	9000	8,9	17	«Villa Arte
13	13	0	0,7223447269	13	2	4000	8,7	2,5	«Корона
14	14	1	0,368645405	14	3	7000	8,5	11,5	«Novik Co
15	15	1	0,7910955682	15	3	4200	9,6	15,6	«Союз»
16	16	0	0,9307202161	16	2	5000	8,5	13,7	Стори
17	17	1	1,189901728	17	2	6500	8,9	0,6	«Сибирс
18	18	0	0,5826427229	18	2	3000	8,3	5,7	«Примор
19	19	1	0,5967682339	19	3	5000	9,3	14,4	«АВРОРА
20	20	0	0,5826427229	20	2	3000	7,9	2,1	«Учебная
21	21	1	1,24688434	21	2	8000	8,9	2,1	Аванта
22	22	0	0,5559889225	22	2	2500	7,6	2,6	Вариант

5. На скриншоте результат оценки линейной регрессии (оценка коэффициентов). Какие входные параметры использованы для операции? Выберите все ответы!

a) Отель
b) Звездность
c) Расстояние до центра
d) Цена
e) Оценка
f) Отель _название



#	Имена входных полей	Метки входных пол…	Уникальные значен…	Цена|Коэффициенты	Цена|Стд. откл.	Цена|T-статисти…	Цена
1	<Константа>			1601,782772	INF	0	
2	COL2	Звёздность	2	0			
3	COL2	Звёздность	3	1845,123992	INF	0	
4	COL2	Звёздность	4	7283,213487	INF	0	
5	COL2	Звёздность	5	13650,67771	INF	0	
6	COL5	Расстояние от центра		193,7233321	INF	0	
7	COL6	Отель_Название	«InCity	84,57923237	INF	0	
8	COL6	Отель_Название	«IZBA»	-779,2720969	INF	0	
9	COL6	Отель_Название	«Novik Country Club»	1325,274903	INF	0	

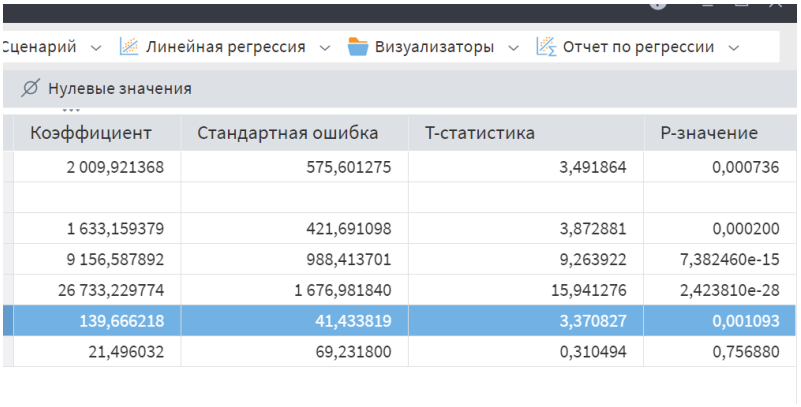
6. Для предыдущей модели укажите какие объясняющее переменные нельзя было включать в модель и почему (Ответ «почему» в развернутом виде)

a) Отель
b) Звездность
c) Расстояние до центра
d) Цена
e) Оценка
f) Отель _название

7. На скриншоте результат оценки линейной регрессии – но здесь заданы другие объясняющие переменные! Какие? Выберите все ответы! Для того чтобы ответить на вопрос нужно выполнить расчеты.

a) Отель
b) Звездность
c) Расстояние до центра
d) Цена
e) Оценка

f) Отель _название



Коэффициент	Стандартная ошибка	T-статистика	P-значение
2 009,921368	575,601275	3,491864	0,000736
1 633,159379	421,691098	3,872881	0,000200
9 156,587892	988,413701	9,263922	7,382460e-15
26 733,229774	1 676,981840	15,941276	2,423810e-28
139,666218	41,433819	3,370827	0,001093
21,496032	69,231800	0,310494	0,756880