



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

Факультет «Информатика и системы управления»
Кафедра «Системы обработки информации и управления»

Домашнее задание
по курсу «Оперативный анализ данных»

Выполнил:
студент группы ИУ5-53Б
Волгина А. Д.
22.11.21

Проверил:
Маслеников К. Ю.

2021 г.

Введение

Датасет “Лаборатория молекулярной биологии” предназначен для решения задач замены неисправного оборудования и выполнения заказов по созданию белков. Сбор данных производится с помощью хранилища данных в Deductor.

Описание предметной области

Предметная область – лаборатория молекулярной биологии. Рассмотрим основные значимые сущности:

- Лаборатория (Laboratory). Атрибуты: ID_lab, ID_medicine, ID_equip, ID_license, ID_employee, Название, Дата_основания, Специализация, Владелец, Годовые_доходы, Годовые_расходы, Количество_филиалов.
- Сотрудники (Employees). Атрибуты: ID_employee, Пол, Возраст, Стаж, Зарплата, Должность.
- Лекарства (Medicines). Атрибуты: ID_medicine, ID_protein, Название, Порция.
- Белки (Proteins). Атрибуты: ID_protein, Название, Вид, Функция, Есть_в_наличии.
- Оборудование лаборатории (Laboratory_equipment). Атрибуты: ID_equip, ID_supplier, Название, Назначение, Исправно.
- Поставщики оборудования (Equipment_sullpiers). Атрибуты: ID_supplier, Название.
- Лицензии на оборудование (Equipment_licenses). Атрибуты: ID_license, ID_company, Название, Оборудование
- Компании-владельцы (Companies_owners). Атрибуты: ID_company, Название, Страна.

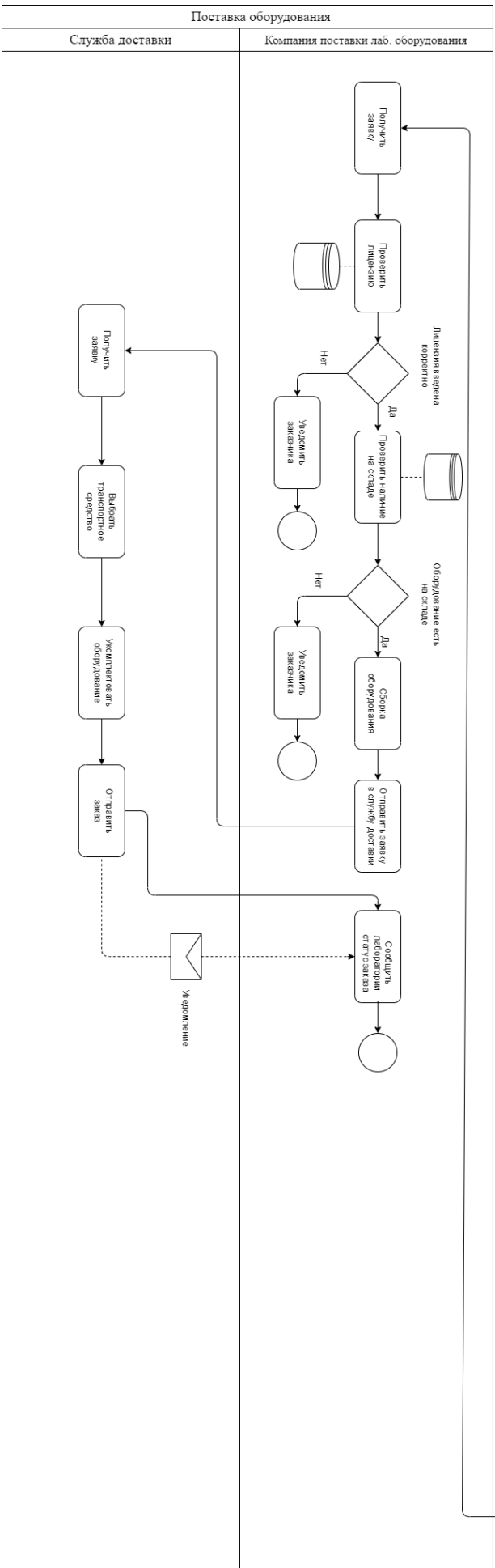
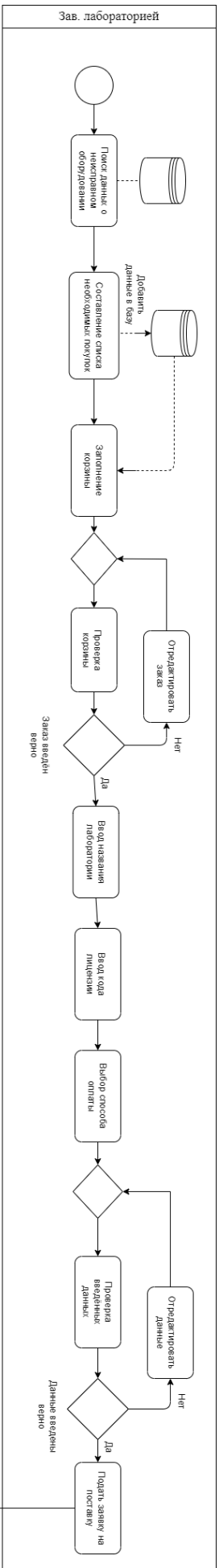
Данное хранилище данных позволяет ответить, например, на следующие вопросы:

- Какое оборудование лаборатории неисправно?
- Какой белок необходим для данного лекарства?

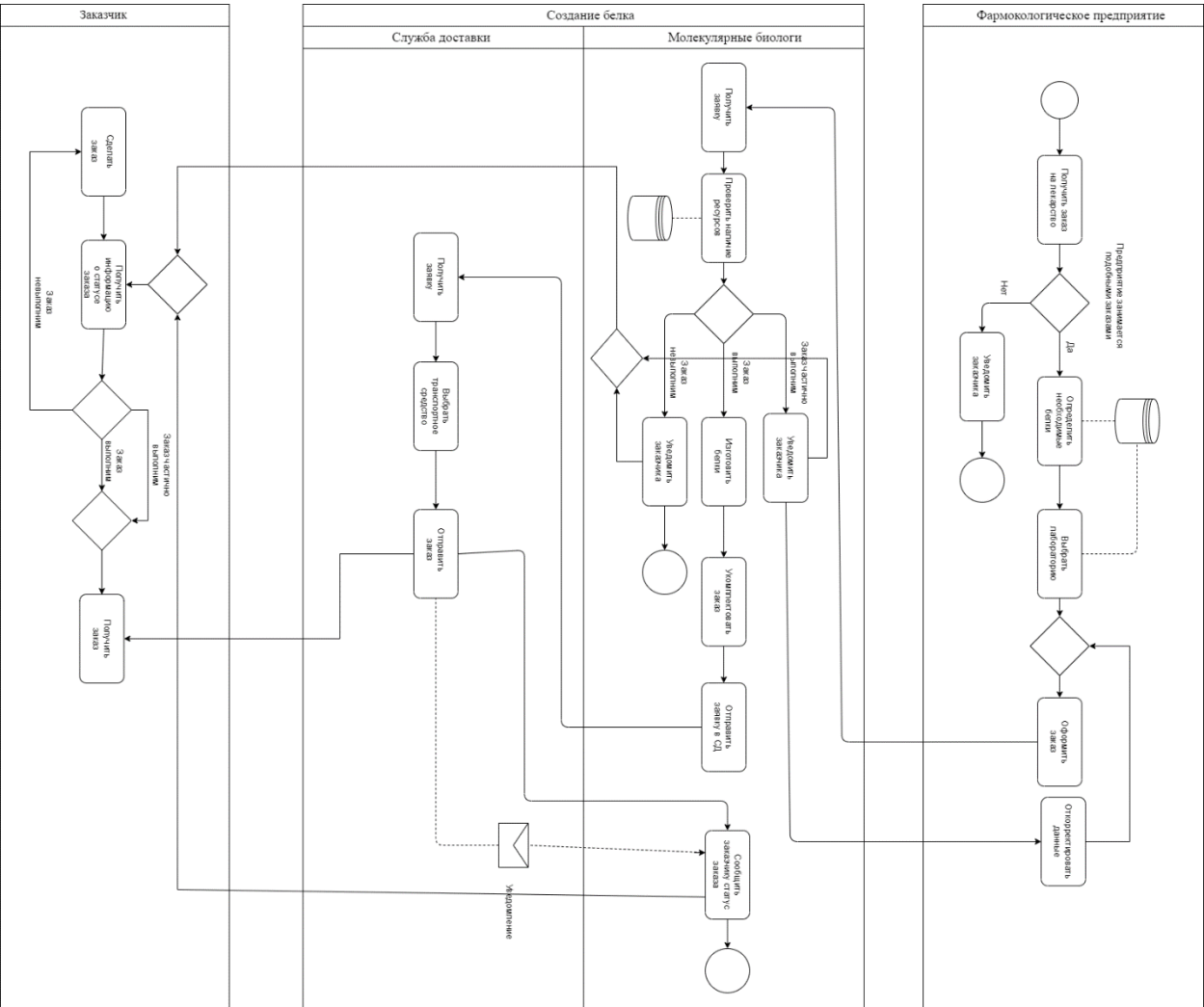
Описание бизнес-процессов в нотации BPMN 2.0

Заведующий лабораторией, обращаясь в хранилище данных, получает информацию о неисправном оборудовании и заказывает новое у компании поставки лаб. оборудования. Компания с помощью хранилища данных проверяет, есть ли у лаборатории лицензия на оборудование, и, если с ней всё хорошо и заказываемый товар есть на складе, то отправляет заявку в службу

доставки. Служба доставки отправляет заказ. Если что-то идёт не так, то заведующему лабораторией отправляют уведомление.

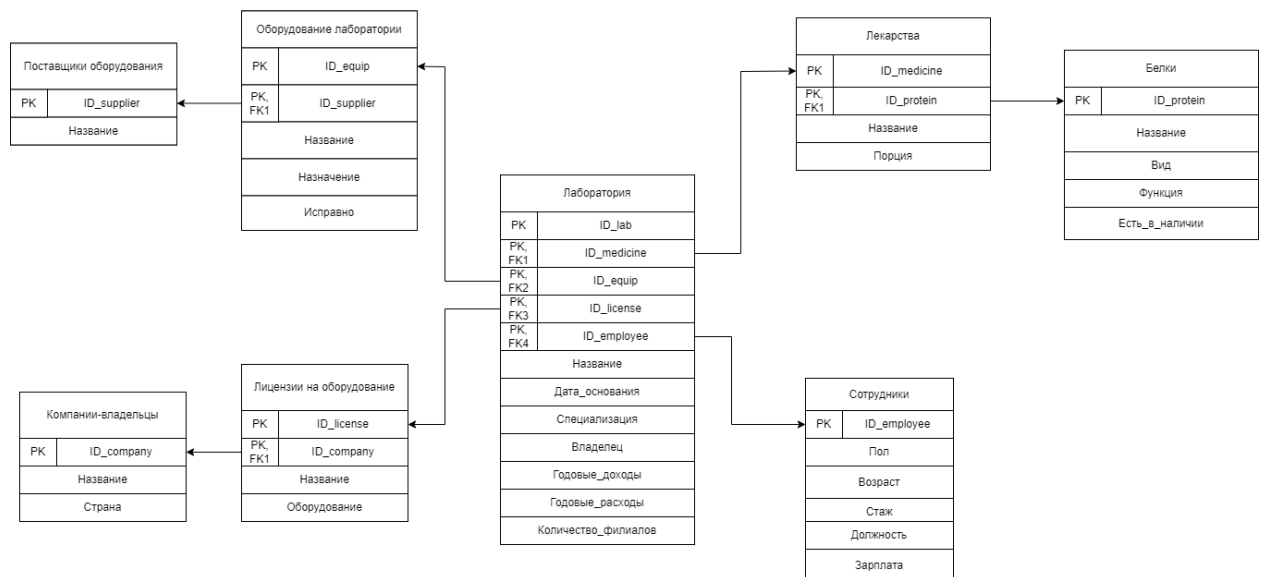


Фармакологическое предприятие получает заказ на лекарство, с помощью хранилища данных определяет необходимые для него белки и заказывает их у лаборатории (молекулярных биологов). Биологи, обращаясь в хранилище данных, определяют, есть ли нужные белки в наличии. Если есть, то отправляют заказ при помощи службы доставки. Если что-то идёт не так, то уведомляют заказчика.

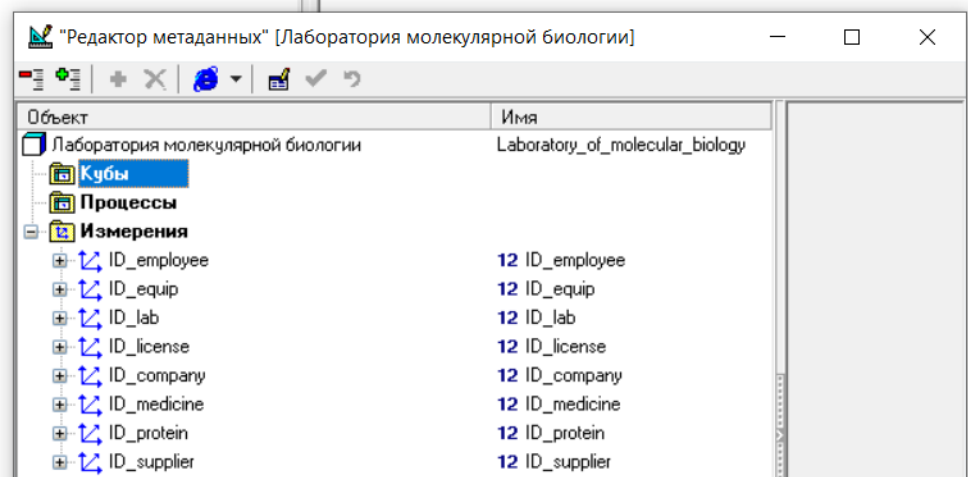
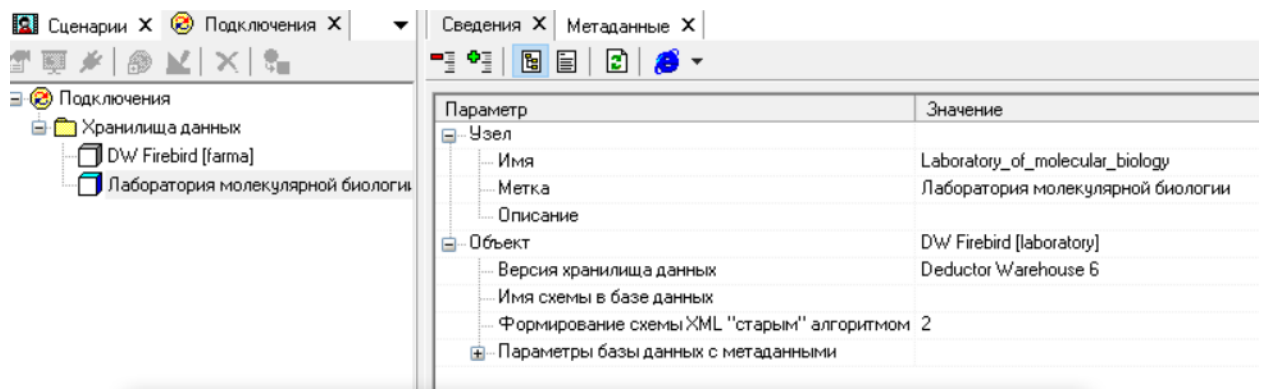


Построение хранилища данных

Нарисуем схему нашего хранилища данных



Построим его в программе Deductor и загрузим в него датасет



Сценарии		ID_company	Название	Страна
Сотрудники	Загрузка данных в Хранилище даннь	1	Свежелов	Китай
Лицензии на оборудование	Загрузка данных в Хранилище даннь	2	Фрешер	Китай
Лекарства	Загрузка данных в Хранилище даннь	3	БонФреш	США
Компании-владельцы	Загрузка данных в Хранилище даннь	4	А-Свеж	США
Белки	Загрузка данных в Хранилище даннь	5	Новичок	Россия
Поставщики оборудования	Загрузка данных в Хранилище даннь	6	Ценовик	Россия
Лаборатория	Загрузка данных в Хранилище даннь	7	Снежанс	Китай
Оборудование лаборатории	Загрузка данных в Хранилище даннь	8	Авоська	Китай
		9	Ажур	Россия
		10	Пользоллов	Россия
		11	Супер-толк	Германия
		12	АльтерМаг	Германия
		13	УмФорма	Россия
		14	Рационик	Россия
		15	Биомаг	США
		16	Кулёк	США
		17	Подкова	Китай
		18	Полюс	Китай

Сценарии		ID_protein	Название	Вид	Функция	Есть_в_наличии
Сотрудники	Загрузка данных в Хранилище даннь	1	гликопротеины	сложный	гормон	<input checked="" type="checkbox"/>
Лицензии на оборудование	Загрузка данных в Хранилище даннь	2	гликопротеины	сложный	гормон	<input type="checkbox"/>
Лекарства	Загрузка данных в Хранилище даннь	3	липопротеины	сложный	транспорт	<input checked="" type="checkbox"/>
Компании-владельцы	Загрузка данных в Хранилище даннь	4	липопротеины	сложный	транспорт	<input type="checkbox"/>
Белки	Загрузка данных в Хранилище даннь	5	фосфопротеины	сложный	регулирование	<input checked="" type="checkbox"/>
Поставщики оборудования	Загрузка данных в Хранилище даннь	6	фосфопротеины	сложный	регулирование	<input type="checkbox"/>
Лаборатория	Загрузка данных в Хранилище даннь	7	гистоны	простой	упаковка нитей ДНК	<input checked="" type="checkbox"/>
Оборудование лаборатории	Загрузка данных в Хранилище даннь	8	гистоны	простой	упаковка нитей ДНК	<input type="checkbox"/>
		9	хромопротеиды	сложный	аккумуляирование энергии	<input checked="" type="checkbox"/>
		10	хромопротеиды	сложный	аккумуляирование энергии	<input type="checkbox"/>
		11	альбумины	простой	траспорт	<input checked="" type="checkbox"/>
		12	альбумины	простой	транспорт	<input type="checkbox"/>

Сценарии		ID_supplier	Название
Сотрудники	Загрузка данных в Хранилище даннь	1	Лавка
Лицензии на оборудование	Загрузка данных в Хранилище даннь	2	Сова
Лекарства	Загрузка данных в Хранилище даннь	3	Скат
Компании-владельцы	Загрузка данных в Хранилище даннь	4	Свекла
Белки	Загрузка данных в Хранилище даннь	5	Дилижанс
Поставщики оборудования	Загрузка данных в Хранилище даннь	6	Лилия
Лаборатория	Загрузка данных в Хранилище даннь	7	Ширма
Оборудование лаборатории	Загрузка данных в Хранилище даннь	8	Ленист
		9	Глобум
		10	Кашира
		11	Нембус
		12	Альферац
		13	Вега
		14	Мега
		15	Ангарес
		16	Хобот
		17	Бульба
		18	БонФреш
		19	Подкова
		20	АмФорма

Таблица										
	ID_equip	ID_supplier	Название				Назначение		Исправно	
	1	14	Радиотомографический комплекс для исследова				Исследования ионосферы и ближнего космоса		<input type="checkbox"/>	
	2	7	Радиотомографический комплекс для исследова				Исследования ионосферы и ближнего космоса		<input checked="" type="checkbox"/>	
	3	11	Система для осаждения алмазных пленок				Осаждение алмазных пленок		<input checked="" type="checkbox"/>	
	4	3	Система для измерения физических свойств PPM				Измерение физических свойств PPMs-9		<input type="checkbox"/>	
	5	17	Гелиевый реконденсатор сверхпроводящего маг				Реконденсация		<input type="checkbox"/>	
	6	4	Когерентный томографический трехчастотный ко				Прием спутниковых сигналов		<input type="checkbox"/>	
	7	20	Многофункциональный сверхширокополосный из				Для исследований свойств нано- метаматериало		<input checked="" type="checkbox"/>	
	8	5	Полуавтоматическая зондовая станция				Прецизионное позиционирование		<input type="checkbox"/>	
	9	14	Исследовательская установка				Нанесение ЛЦУ пленок и различных гетерострук		<input type="checkbox"/>	
	10	20	Когерентный томографический трехчастотный ко				Прием спутниковых сигналов		<input checked="" type="checkbox"/>	
	11	1	Когерентный томографический трехчастотный ко				Прием спутниковых сигналов		<input checked="" type="checkbox"/>	
	12	2	Мессбауэровский спектрометр				Структурный анализ		<input type="checkbox"/>	
	13	9	Когерентный томографический трехчастотный ко				Прием спутниковых сигналов		<input type="checkbox"/>	
	14	5	Гелиевый реконденсатор сверхпроводящего маг				Реконденсация		<input type="checkbox"/>	
	15	2	Многофункциональный сверхширокополосный из				Для исследований свойств нано- метаматериало		<input checked="" type="checkbox"/>	
	16	19	Мессбауэровский спектрометр				Структурный анализ		<input type="checkbox"/>	
	17	11	Гелиевый реконденсатор сверхпроводящего маг				Реконденсация		<input type="checkbox"/>	
	18	6	Исследовательская установка				Нанесение ЛЦУ пленок и различных гетерострук		<input checked="" type="checkbox"/>	
	19	19	Полуавтоматическая зондовая станция				Позиционное позиционирование		<input type="checkbox"/>	

	ID_lab	ID_medicine	ID_equip	ID_license	ID_employee	Название	Дата_основания	Специализация	Владелец	Годовые_доходы	Годовые_расходы
	1	776	423	161	1084	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	549	240	195	360	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	841	4	317	174	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	504	380	3	644	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	18	182	475	474	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	499	40	51	190	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	904	409	385	271	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	895	272	413	509	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	281	212	187	1040	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	466	258	432	654	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	754	404	377	496	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	1030	448	360	995	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	627	56	329	1112	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	1057	167	213	384	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	247	143	58	985	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	667	318	408	309	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	1009	487	281	973	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	15	327	485	1078	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	236	265	251	300	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	847	249	261	802	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	125	331	349	513	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	778	425	172	707	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	130	2	296	944	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	93	278	45	175	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	765	47	438	437	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	793	58	305	387	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	1157	143	65	85	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	478	124	186	717	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000
	1	1051	147	43	597	BostonGene	04.05.1999	Молекулярная биология	Мария Добычина	200000	50000

Визуализация данных

Визуализируем наш датасет в Tableau.

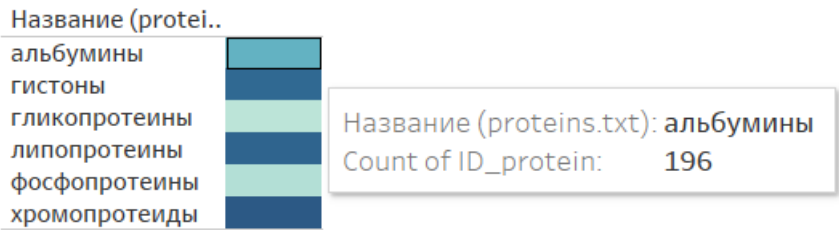
Посмотрим количество неисправного оборудования:

Количество неисправного оборудования

Название (equipment.txt)		
Электронно-зондовый микроанализатор JXA-8230		20
Гелиевый реконденсатор сверхпроводящего магнита		15
Исследовательская установка		18
Когерентный томографический трехчастотный комплекс приема спутниковых сигнал..		24
Конфокальный КР микроскоп		23
Мессбауэровский спектрометр		22
Многофункциональный сверхширокополосный измерительный комплекс параметро..		16
Полуавтоматическая зондовая станция		26
Радиотомографический комплекс для исследований ионосферы и ближнего космоса		28
Система для измерения физических свойств PPMs-9		25
Система для осаждения алмазных пленок		25

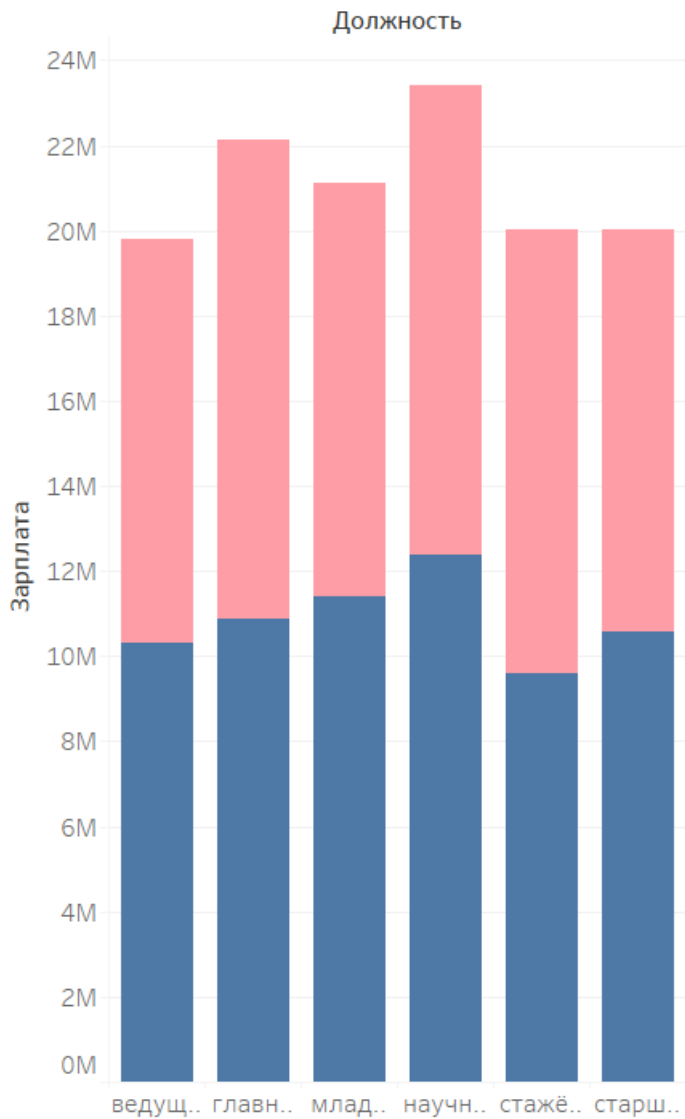
Количество используемых в лекарствах белков

Количество используемых белков



Распределение денег на зарплату сотрудников по полу (считается общая сумма для зарплат работникам по данной должности)

Распределение денег на зарплату сотрудников по полу



Количество используемых белков

Название (protei..

альбумины	
гистоны	
гликопротеины	
липопротеины	
фосфопротеины	
хромопротеиды	

Count of ID_protein

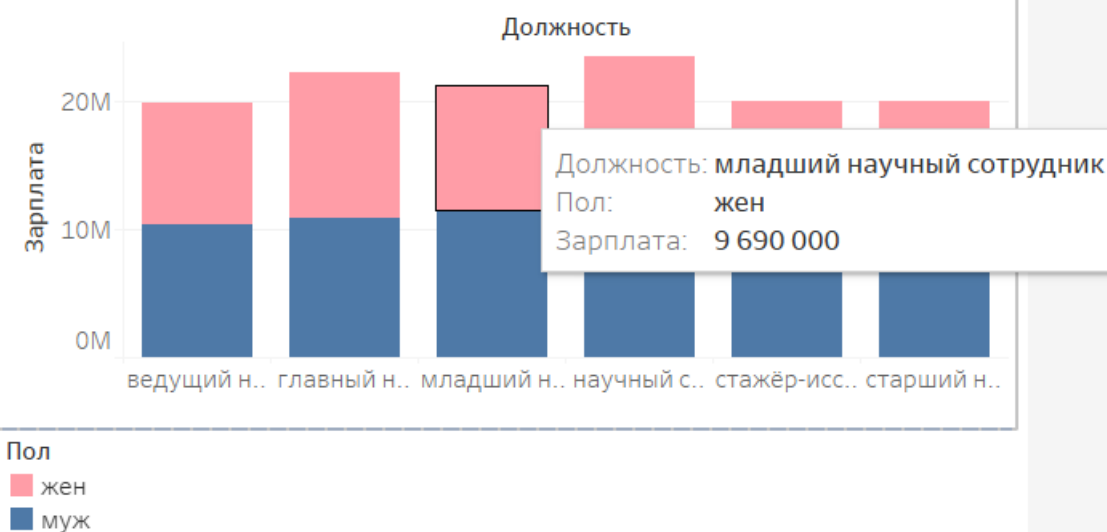
186 212

Количество неисправного оборудования

Название (equipment.txt)

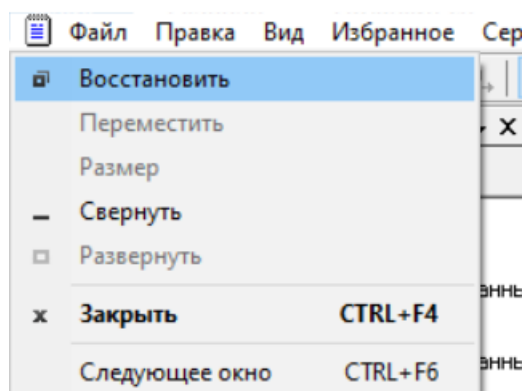
Электронно-зондовый микроанализатор
Гелиевый реконденсатор сверхпроводящего магнита
Исследовательская установка
Когерентный томографический трехчастотный источник света
Конфокальный КР микроскоп
Мессбауэровский спектрометр
Многофункциональный сверхширокополосный усилитель
Полуавтоматическая зондовая станция
Радиотомографический комплекс для исследования биологических объектов
Система для измерения физических свойств материалов
Система для осаждения алмазных пленок

Распределение денег на зарплату сотрудникам по полу



План архивации и восстановления данных

Восстановить данные можно с помощью интерфейса программы



Заключение

В результате выполнения домашнего задания я ознакомилась с предметной областью лаборатории молекулярной биологии, описала бизнес-процессы в нотации BPMN 2.0, создала хранилище данных в Deductor и несколько визуализаций при помощи Tableau.

Литература

- Документация по Deductor <https://basegroup.ru/deductor/function>
- Документация по Tableau <https://www.tableau.com/learn>