

## Лабораторная работа № 9

### Тема: «Проведение анализа вакансий. Моделирование данных в Power Bi»

**Цель работы:** получить навыки моделирования данных в Power Bi на примере данных с вакансиями.

#### Теоретическая справка

##### Текстовые функции DAX

Таблицы и столбцы можно использовать с текстовыми или строковыми функциями DAX. Вы можете объединять строковые значения, искать текст со строкой и возвращать часть строки с помощью текстовых функций DAX. Форматы для дат, времени и чисел также настраиваются

.No	DAX – Текстовые функции	Описание
1	BLANK ( )	Возвращает пробел.
2	LEN ( )	Возвращает количество символов в текстовой строке.
3	CONCATENATE ( )	Объединяет две текстовые строки в одну текстовую строку.
4	CONCATENATEX ( )	Объединяет результат выражения, вычисленного для каждой строки в таблице.
5	LEFT ( )	Возвращает указанное количество символов с левой стороны заданной текстовой строки.
6	RIGHT ( )	Возвращает указанное количество символов с правой стороны заданной текстовой строки.
7	MID ( )	Возвращает строку символов из середины текстовой строки.
8	UPPER ( )	Преобразует текстовую строку в заглавные буквы.
9	LOWER ( )	Преобразует текстовую строку в строчные буквы.
10	TRIM ( )	Удаляет все пробелы из текста.

## «Визуальная аналитика», Киреев В.С.

No	DAX – Текстовые функции	Описание
11	SUBSTITUTE ( )	Заменить существующий текст новым текстом в текстовой строке.
12	REPLACE ( )	Заменяет часть текстовой строки.
13	EXACT ( )	Сравнивает две текстовые строки и возвращает TRUE, если они абсолютно одинаковы, в противном случае возвращает FALSE.
14	FIND ( )	Возвращает начальную позицию одной текстовой строки внутри другой текстовой строки.
15	FORMAT ( )	Преобразует значение в текст в соответствии с указанным форматом.

Функция Find возвращает начало одной текстовой строки внутри другой текстовой строки.

Синтаксис: FIND(<Find\_Text>, <Within\_Text>, [ <Start\_Num>], [<NotFoundValue>])

- Любой текст или строка, которые вы найдете с помощью функции find text, — текст, содержащий искомый текст.
- StartPosition: не обязательно. Начальной точкой поиска должна быть позиция символа. Изначально это 1. NotFoundValue использовать необязательно. Обычно 0, -1 или BLANK, это значение должно быть возвращено, если операция не может определить соответствующую подстроку ().


### Логические функции

SWITCH - возвращает одно из нескольких потенциальных результирующих выражений после сравнения выражения со списком значений.

Синтаксис: SWITCH(<выражение>, <значение>, <результат>[, <значение>, <результат>]...[, <иначе>])

Country	Product	Qty Sold	Year of Sales	Gross Sales	Net Sales	NotDemo	CountryCode
Canada	Pen	20	2019	420	400	Invalid	CA
Germany	Mouse	15	2019	30	15	Valid	GE
France	CPU	1999	2018	9995	5997	Valid	FR
Germany	Pen	350	2019	2450	350	Valid	GE
Mexico	CPU	12	2018	156	96	Valid	
Germany	Mouse	7	2020	84	70	Valid	GE
Germany	Pencil	15	2019	45	15	Valid	GE
Canada	Pencil	120	2020	5280	4800	Invalid	CA
France	Pen	15	2019	315	165	Valid	FR
Germany	CPU	7	2018	238	140	Valid	GE
Mexico	Mouse	22	2020	462	242	Valid	
Canada	Mouse	12	2019	672	276	Invalid	CA
Mexico	Mouse	15	2019	825	330	Valid	
Germany	Mouse	20	2019	1540	660	Valid	GE
Canada	Pen	800	2018	71200	64800	Invalid	CA
United States of America	Pen	7	2019	63	21	Valid	
Canada	Pen	125	2020	9500	9375	Invalid	CA


## «Визуальная аналитика», Киреев В.С.

Для структурированного столбца, такого как List, Record, или Table, есть связанные значения, связанные со столбцом. Структурированные столбцы имеют значок Expand  в заголовке столбца.

### Структурированные колонки

- **Список** Хранит список значений, например, { 1, 2, 5, 10 } или { "привет", "привет", "до свидания" }. Значение списка в ячейке является простейшим структурированным столбцом, и нет никакой табличной связи.
- **Record** Хранит набор именованных полей, сгруппированных в блок. Например, [FirstName = "Joe", LastName = "Smith", Birthdate = #date(2010, 1, 2)]. Значение Record в ячейке указывает на связь таблицы "один к одному" с текущей или первичной таблицей.
- **Таблица** Хранит таблицу, имеющую вторичную связь с текущими данными, которые являются первичной таблицей. Значение Таблица в ячейке указывает на связь «один ко многим» с текущей или первичной таблицей.

Структурированный столбец «Запись и таблица» представляет собой связь в источнике данных, имеющую реляционную модель, например, сущность с ассоциацией внешнего ключа в канале OData или связь внешнего ключа в базе данных SQL Server.

Вы можете использовать функцию «Расширить»  значок для просмотра столбцов из связанной таблицы. Например, в таблице Orders операция расширения объединяет записи Order\_Details, связанные с таблицей Order, для объединения позиций заказа с каждым заказом. Операция расширения расширяет первичную таблицу, включая столбцы из вторичной и связанной таблиц.

Первичная таблица имеет столбцы A и B.

A	B

Связанная таблица имеет столбец C.

A	B	C

C

Операция расширения расширяет первичную таблицу, включая столбец C, и расширяет первичную таблицу связанными значениями из вторичной и связанной таблицы, содержащей столбец C.

A	B	C

## «Визуальная аналитика», Киреев В.С.

### Самостоятельное задание

1. В качестве источника данных для анализа вакансий, используйте данные парсинга сайта hh.ru по запросу «Аналитик», в 4-х городах (Москва, Санкт-Петербург, Екатеринбург, Новосибирск).
2. Создайте вычисляемый столбец – «Зарплата» (среднее значение по столбцам “salary.from”, “salary.to”, умноженное на коэффициент 1 (“salary.gross” = False) или 0.87 (“salary.gross” = True))
3. Создайте вычисляемый столбец – «Квалификация» («Сеньор» (“experience.name ” = “Более 6 лет” или “От 3 до 6 лет”) и (“name” содержит “Senior” или “Ведущий” или “Старший”)), «Миддл» (“experience.name ” = “От 1 года до 3 лет”), «Джун» (“experience.name ” = “Нет опыта”) и (“name” содержит “начинающий” или “стажер” или “помощник”), «Другое», содержащее все остальные вакансии)
4. Создайте по аналогии с предыдущими заданиями, ориентируясь на поле «name», вычисляемый столбец – «Область» («дата-аналитик», «BI-аналитик», «системный аналитик», «бизнес-аналитик», «web-аналитик», «финансовый аналитик» и «Другое»)
5. Создайте по аналогии с предыдущими заданиями, ориентируясь на поля «name», «snippet.requirement», «snippet.responsibility», вычисляемый столбец – «Инструменты» («Power BI», «Python», «Excel», «1C», «DataLens», «Tableau», «Apache Superset» и «Другое»)
6. Создайте по аналогии с предыдущими заданиями, ориентируясь на поля «name», «snippet.requirement», «snippet.responsibility», вычисляемый столбец – «Навыки» («BPMN», «UML», «ARIS» и «Другое»)
7. Создайте графики распределений:
  - 7.1. Гистограмма распределения зарплат
  - 7.2. Ящичковая диаграмма распределения зарплат внутри городов
  - 7.3. Ящичковая диаграмма распределения числа вакансий внутри типов графика работы (удаленная работа, полный день и пр)
8. Создайте иерархию - Область, Квалификация, Навыки. Создайте меры внутри иерархии – Средняя зарплата, Медианная зарплата. Создайте на их основе графический элемент - Таблицу, содержащую число вакансий и зп (среднюю и медианную), внутри иерархии.
9. Соберите из элементов, созданных в предыдущих заданиях дэшборд похожего вида:

# «Визуальная аналитика», Киреев В.С.

## Анализ рынка вакансий Аналитики и BI

Проект: Reveal The Data и Left Join | Данные с портала hh.ru — методика исследования

Тип ключевого слова: (All) | Города: (All) | Среднее/Медиана: Среднее | Расчет зарплаты по: средней в отрасли | Дата публикации: Last 3 months

### Вакансии по типам навыка

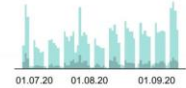
Выберите ключевое слово или раздел, чтобы отфильтровать другие графики и таблицы

Раздел	Ключевое слово	Вакансии	01.07.20	10.09.20	Зарплата	0 P	200 000 P	400 000 P	600 000 P
Профессия	Analytics	Бизнес аналитик OR Business Analyst	3 876		82 092 P				
		Аналитик данных OR Data analyst	615		63 078 P				
		Продуктовый аналитик OR Product Analyst	414		90 485 P				
		Аналитик BI OR BI Analyst	173		90 395 P				
		Руководитель отдела аналитики OR Head of ..	89		110 665 P				
		Аналитик баз данных	50		59 452 P				
		Marketing Analyst или маркетинговый аналит...	2						
Профессия	BI	Разработчик BI OR BI Developer	166		111 742 P				
		Инженер баз данных OR DWH Engineer OR ..	91		154 575 P				
		Инженер баз данных OR DWH Engineer	3						
Профессия	Data science	Data scientist OR Machine Learning speci..	222		244 463 P				
		Руководитель направления предикти..	120		135 250 P				
		Руководитель data science	116		143 343 P				
		Data Science Director OR Директор по боль..	3		250 000 P				
Навык	Analytics	Python AND data AND анализ	490		128 022 P				
		Pandas	405		83 519 P				
		Matplotlib	86		85 350 P				
		NumPy	67		157 407 P				
		Plotly	31		78 433 P				
	BI	Power BI	632		72 095 P				
		Дашборд	559		96 861 P				

### Вакансий

11875

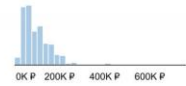
2137 с зарплатой + 1676 за семь дней



### Зарплата

96 096 P

max — 700 000 P  
min — 139 P



### По типу занятости

Выберите строку для фильтрации

Полный день	9 891	95 861 P
Удаленка	1 179	99 255 P
Гибкий график	731	89 511 P

### По городам

Выберите город для фильтрации

Москва	5 895	121 241 P
Санкт-Петербург	1 507	102 197 P
Минск	528	43 750 P
Новосибирск	321	87 278 P
Краснодар	273	70 976 P
Екатеринбург	271	65 437 P
Киев	270	
Казань	216	66 532 P

10. Добавьте фильтры «Город», «График работы».