Тестовое задание на позицию

Data Analyst

Данное тестовое состоит из нескольких подзаданий. В идеальном варианте мы ожидаем выполнение всех подзаданий. Наряду с непосредственно расчетами важны подача результата и изложение Вашего хода мысли. Ограничений по использованию инструментария и форме финального отчета нет.

Результаты тестового (архив) и Ваши комментарии к нему (сколько времени суммарно было затрачено (в часах), какие сложности были, какая задача понравилась больше всего, почему и т.п.) необходимо выслать на e-mail.

Желаем успеха!

Задание 1

В файле «Publishing\_Analyst\_Data.xls» на вкладке **«Основные характеристики»** имеется некоторая выборка. Вам необходимо

* построить дискретный вариационный ряд и вычислить следующие основные характеристики:
  + среднее значение
  + моду
  + медиану
  + дисперсию
  + среднеквадратическое отклонение
  + размах
  + коэффициент ассимметрии
  + коэффициент эксцесса
* построить полигон частот и кумулятивную кривую
* пояснить, что означает каждая из характеристик и сделать выводы.

Задание 2

В файле «Publishing\_Analyst\_Data.xls» на вкладках «**Для определения зависимости 1**» и «**Для определения зависимости 2**» имеются две выборки: «Выборка 1» и «Выборка 2». Ваша задача – определить степень зависимости между ними, сделать выводы, пояснить выбранную методику расчета.

Задание 3

Зависимость между переменной X и Y имеет вид y=**Ax**+B+e, где ‘e’ – случайные ошибки. Построить линейную регрессионную модель и оценить параметры **A** и B. Указать применяемый метод, проверить адекватность полученной модели и дать пояснения полученным результатам. Данные находятся в файле «Publishing\_Analyst\_Data.xls» на вкладке **«Построение регрессии».**

Задание 4

В «Минсктранс» поступила жалоба от пассажиров на плохую работу транспорта на диспетчерской станции «Дружная» в период с 20:00 до 21:00. Суть жалобы состоит в том, что троллейбусы ходят не по расписанию и их приходится ждать очень долго на остановках. На предприятии «Минсктранс» существует система регистрации времени прибытия троллейбусов, а также подсчета количества пассажиров. Данные с системы регистрации приведены в файле «Publishing\_Analyst\_Data.xls» на вкладке **«Статистика по троллейбусам»**. Руководство предприятия поручило вам разобраться в ситуации и выяснить, чем же вызвана жалоба, а также провести развернутый анализ времени движения транспорта и пассажиропотока.

Требуется

* найти причину жалобы,
* выяснить характеристики, которыми характеризуется движение транспорта и пассажиропоток (числовые, графические, как по отдельным маршрутам троллейбусов, так и в совокупности)
* предоставить развернутую характеристику полученных результатов.

*Постарайтесь писать «отчет» наиболее понятным для несведущего в статистике руководства предприятия языком. Например, не рекомендуется писать, что математическое ожидание движения троллейбусов равно 10, а лучше написать — средний интервал движения троллейбусов равен 10 минутам.*

Задание 5

Существует четыре рекламных канала регистраций пользователей A, B, C и D. В сентябре канал D был удален, и по май включительно существовало только три рекламных канала. В июне предполагается снова ввести рекламный канал D. Необходимо определить будущее значение регистраций на нем в июне. Количество регистраций по периодам на каждом канале представлено в файле «Publishing\_Analyst\_Data.xls» на вкладке **«Рекламные каналы»**.

Задание 6

Составить аналитический отчет о продажах фруктов. Данные для составления отчета приведены в файле «Publishing\_Analyst\_Data.xls» на вкладке **«Продажи фруктов»** в килограммах.

Задание 7

Существует несколько каналов привлечения новых пользователей (регистраций). Они могут самостоятельно набирать в браузере необходимую ссылку и затем регистрироваться (organic), либо они могут привлекаться с помощью рекламных кампаний (adv, например, пользователю показывается баннер, ведущий на сайт регистрации, для данных пользователей выставляется отметка adv). Так же запускаются многочисленные ТВ кампании, в результате проведения которых, к сожалению, нет возможности выставить пользователям отметку, что он зарегистрировался в результате просмотра ТВ-рекламы.

С 21.06.2015 по 11.07.2015 проводилась рекламная ТВ-кампания по привлечению новых пользователей в World of Tanks. Количество регистраций по дням находится в файле «Publishing\_Analyst\_Data.xls» на вкладке **«Data for TV Analysis»**, колонка «Total».

Динамика количества регистраций пользователей, которые пришли самостоятельно расположены в колонке “Organic”, по рекламным каналам – в колонке “Adv”.

Известно, что среднестатистический пользователь приносит проекту 10$ за всю свою жизнь. Так же известно, что компания тратит на привлечение пользователя по рекламным каналам 1$.

Посчитать потенциальную прибыль от рекламной кампании, если затраты на ее проведение составили 100 000$.

Задание 8

Оцените объем целевой аудитории игры World of Tanks в Республике Молдова.

*----------------------------------------------------------------------------------------------------------------------------------------------*

*Отчет должен содержать:*

* *Описание методики исследований и используемые алгоритмы*
* *Найденные эффекты с пояснениями*
* *Результаты анализа отчета*

*Рекомендации и комментарии всего тестового задания:*

* *можно использовать любые подручные средства и любые пакеты*
* *приводите применяемые методы и обосновывайте полученные результаты*
* *если по каким-то причинам все не получится, присылайте то, что получилось*
* *желательная форма отчета – презентация, содержащая основные выводы, описания методик и наблюдения, составленная в максимально доступном и понятном ключе + набор приложений с детальным описанием решения каждого из заданий*