Темы для проведения опроса

1. Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.
2. Этапы анализа данных.
3. Структурированные и неструктурированные данные. Сбор и подготовка данных.
4. Библиотека pandas.
5. Структуры данных в pandas, работа со структурами данных.
6. Операции над данными.
7. Комбинирование данных из разных источников.
8. Обработка пропущенных значений.
9. Виды данных и их загрузка.
10. Структурированные и неструктурированные типы данных. Основные форматы файлов - txt, csv, xls, sav.
11. Структура файлов.
12. Виды разделителей, символы окончания строки.
13. Проблема кодировок и различия операционных систем. SQL-базы данных. Удаленные базы данных (API).
14. Неструктурированные данные - json, xml.
15. Сохранение или запись файлов, представление в внешних веб-приложения
16. Введение в визуализацию данных.
17. Нюансы визуализации данных и принципы человеческого восприятия.
18. Правила создания хороших визуализаций.
19. Обзор различных видов графиков (гистограмма, бар-чарт, секторная диаграмма, линейные графики, график рассеяния, тепловая карта и т.д.).
20. Особенности разных видов графиков и их использования.
21. Задачи визуализации данных.
22. Статичные графики, интерактивные визуализации, инфографика.
23. Виды графиков.
24. Принципы визуальной презентации данных.
25. Ошибки в использовании линейных графиков, гистограмм, круговых и объемных диаграмм. Палитры для графиков.
26. Статистика: описание одиночного набора данных, показатели центра распределения, показатели вариации, корреляция и причинная зависимость.
27. Проверка статистических гипотез.
28. Оценка выборочных параметров с использованием специальных функций.
29. Законы распределения вероятностей в R.
30. Проверка на нормальность распределения.
31. Классические методы и критерии статистики.
32. Гипотеза о равенстве средних двух генеральных совокупностей.
33. Рандомизация, бутстрап и оценка статистической мощности. Введение в дисперсионный анализ.
34. Оценка корреляций двух случайных величин. Критерий хи-квадрат.
35. Регрессионные модели зависимостей между количественными переменными.
36. Модель простой линейной регрессии.
37. Модели регрессии при различных видах функции потерь.
38. Полиномиальные и нелинейные модели регрессии.
39. Сравнение эффективности различных моделей при прогнозировании.
40. Модели для оценки временных рядов.
41. Определение стационарности временных рядов.
42. Определение ложных регрессий.
43. Работа с трендами.
44. Гармонический и спектральный анализ.
45. Построение и прогнозирование авторегрессионных моделей ARIMA.
46. Модели ADL, ARCH, GARCH.
47. Многомерные методы анализа данных