

Проектное предложение

Тип элемента практической подготовки	Проект
Если проект, тип проекта	Исследовательский
Наименование проекта	Обработка данных для расчета ИПЦ в России с применением «больших данных»
Подразделение инициатор проекта	Институт государственного и муниципального управления / Центр экономической экспертизы
Руководитель проекта	Гуленков Илья Васильевич
Основное место работы руководителя проекта в НИУ ВШЭ	Институт государственного и муниципального управления / Центр экономической экспертизы
Контакты руководителя (адрес эл. почты)	iluagulenkov@gmail.com
Соруководители проекта от НИУ ВШЭ (если имеются)	
Контакты соруководителей от НИУ ВШЭ (адрес эл. почты)	
Основная проектная идея / описание решаемой проблемы	<p>Индекс потребительских цен – один из ключевых макроэкономических параметров. Методология расчета ИПЦ Росстата предполагает отслеживание цен напрямую в оффлайн-магазинах. При этом параллельно во многих странах активно обсуждается применение «больших данных» – множества ценовых котировок, собранных в Интернете – для уточнения данных ИПЦ.</p> <p>Проект направлен на систематизацию международного опыта таких расчетов, а также решение прикладных задач, возникающих при анализе динамики потребительских цен с применением «больших данных»: классификации произвольного товара в одну из рубрик классификатора (КИПЦ), отсева систематических и случайных ошибок, методов построения отдельных индексов цен на основе собранных данных.</p>
Цель и задачи проекта	Первичная подготовка данных о выставленных на продажу в

	<p>российских розничных онлайн-магазинах потребительских товарах и их ценах.</p>
<p>Проектное задание</p>	<p>Каждый из участников получает индивидуальное задание по анализу определенной группы товаров из классификатора КИПЦ (классификатора индивидуального потребления по целям).</p> <p>Задания предполагают:</p> <ol style="list-style-type: none"> 1) Изучение общей методологии Росстата в части расчёта индекса потребительских цен (ИПЦ) 2) Определение описательных характеристик конкретных товаров и услуг, относящихся к соответствующим категориям КИПЦ 3) Ознакомление с международным опытом применения «больших данных» при расчете ИПЦ 4) Отбор товаров из базы данных (есть в наличии), соответствующих рассматриваемым категориям КИПЦ 5) Определение набора ключевых слов, характеризующих данную группу товаров (с применением метрик TF-IDF и пр.) 6) Подготовка аналитической записки с описанием основных результатов, полученных в ходе работы
<p>Планируемые результаты проекта, специальные или функциональные требования к результату</p>	<p>Размеченная база данных товаров, соотносящая их названия с категориями КИПЦ</p>
<p>Дата начала проекта</p>	<p>01.07.2022</p>
<p>Дата окончания проекта</p>	<p>31.08.2022</p>
<p>Трудоемкость (часы в неделю) на одного участника</p>	<p>8</p>

Предполагаемое количество участников (вакантных мест) в проектной команде	5
<p>Названия вакансий (ролей), краткое описание задач по каждой вакансии, количество кредитов и критерии отбора для участников проекта (<i>если характер работ для всех участников совпадает, описывается одна вакансия</i>)</p> <p><i>Кредиты на 1 участника рассчитываются по формуле: продолжительность в неделях * трудоемкость проекта в часах / 25</i></p>	<p>Роли всех участников проекта одинаковы. Для успешного выполнения задач от студента требуется умение работать с текстовой информацией и базовые навыки программирования (Python/R). Число кредитов на одного участника: $9 \cdot 8 / 25 = 2.88 \Rightarrow 3$ кредита</p>
Общее количество кредитов	15 кредитов
Форма итогового контроля	Экзамен в форме защиты итоговых результатов
Формат представления результатов, который подлежит оцениванию	Размеченная база данных + краткая аналитическая записка по итогам исследования
Формула оценки результатов, возможные критерии оценивания результатов с указанием всех требований и параметров	<p>$O_{\text{итог}} = 0.7 \cdot O_{\text{база}} + 0.3 \cdot O_{\text{записка}}$</p> <p>$O_{\text{база}}$ – оценка качества размеченной базы товаров. Основным критерий – безошибочность разметки, а также число отобранных позиций.</p> <p>$O_{\text{записка}}$ – оценка качества аналитической записки. Критерии: самостоятельность, логика и структура изложения, грамотность работы со статистическими методами, качество выводов.</p>
Возможность пересдач при получении неудовлетворительной оценки	Да
Ожидаемые образовательные результаты проекта	В ходе проекта студенты знакомятся с методологией расчета индекса потребительских цен в России, а также основными методами статистического анализа текстовых

	данных, совершенствуют навыки программирования в среде Python/R
Особенности реализации проекта: территория, время, информационные ресурсы и т.п.	Мск, удаленная работа
Рекомендуемые образовательные программы	Мировая экономика
Требуется резюме студента	Нет
Требуется мотивированное письмо студента	Да